OPEN

Correspondence and
requests for materials
should be addressed to
G.C.V. (grish@imtech.
res.in) or G.P.S.R.
(raghava@imtech.res.
in)

* These authors
contributed equally to
this work.

Herceptin Resistance Database for Understanding Mechanism of Resistance in Breast Cancer Patients

Sahil Ahmad*, Sudheer Gupta*, Rahul Kumar, Grish C. Varshney & Gajendra P. S. Raghava

CSIR-Institute of Microbial Technology Chandigarh, India.

Monoclonal antibody Trastuzumab/Herceptin is considered as frontline therapy for Her2-positive breast cancer patients. However, it is not effective against several patients due to acquired or *de novo* resistance. In last one decade, several assays have been performed to understand the mechanism of Herceptin resistance with/without supplementary drugs. This manuscript describes a database HerceptinR, developed for understanding the mechanism of resistance at genetic level. HerceptinR maintains information about 2500 assays performed against various breast cancer cell lines (BCCs), for improving sensitivity of Herceptin with or without supplementary drugs. In order to understand Herceptin resistance at genetic level, we integrated genomic data of BCCs that include expression, mutations and copy number variations in different cell lines. HerceptinR will play a vital role in i) designing biomarkers to identify patients eligible for Herceptin treatment and ii) identification of appropriate supplementary drug for a particular patient. HerceptinR is available at http://crdd.osdd.net/raghava/herceptinr/.

Among targeted therapies in oncology, monoclonal antibodies (mAbs) based therapy is one of the most successful strategies. Herceptin, a recombinant humanized monoclonal antibody targeted against the extracellular domain (ECD) of the HER2 protein[1], ranks among the most significant advances in breast cancer therapeutics[2]. Upon binding to its cognate epitope, Herceptin exerts its antitumor effects by a variety of proposed mechanisms[3]. However, despite this noteworthy attainment, 70% of patients with HER2-positive breast cancers do not get the benefit because of *de novo* or acquired resistance to Herceptin[4]. In this regard, general medical practice exploits various biomarkers to identify patients eligible for treatment with Herceptin[5–7]. This strategy not only renders a cost effective medication but also suggests medical practitioners to change the drug as per patient's constraint. Unfortunately, reliability of available Herceptin biomarkers (diagnostic tests) is very poor[5,8,9]. With the advent of technology particularly high throughput sequencing technologies, it is possible to design genome-based biomarkers for personalized therapy (the right drug for the right patient)[10]. These genome-based biomarkers may utilize expression, mutation or copy number variations of certain genes[11]. In case of Herceptin, various diagnostic kits are available which exploits various molecular-biology techniques to detect amplification/expression of HER2 gene/protein[12,13]. This in turn shows the primitive and underdeveloped form of diagnostics. In order to understand the mechanisms and factors involved in Herceptin resistance, various studies have been performed in the past. However, these studies have been done on different platforms, with tumor tissue samples and cell lines, and taking different aspects like Herceptin response, mutational, expression and copy number variation (CNV) in related genes, effect of supplementary drugs etc. Based on this inhomogeneous scattered data, a gross view with conclusive remarks cannot be made. Thus, it becomes imperative to collect information regarding response of Herceptin, genomic factors causing resistance and probable supplementary drug combination.

In this study, we have made systematic attempts to collect and compile data from various resources to develop a comprehensive database on Herceptin Resistance. This database contains information about 2500 assays, 30 cell lines and 100 supplementary drugs. In order to facilitate researchers, numerous user-friendly tools have been integrated that includes searching, browsing and alignment of genomic data.

Database description and utility

Assay data. This section includes the exploration of experiments performed with Herceptin antibody on different BCCs. The assay data includes experimental details in the form of antibody (Ab) amount, time of Ab treatment (in vitro) supplementary drug, drug amount, time of drug treatment (in vitro), % -inhibition, experimental

techniques and testing Herceptin resistance with cell lines having defined alterations. Our web server provides two major options to explore the data:

Search. This option is meant to search particular keyword such as name of cell line, supplementary drug, status in terms of resistance or sensitive, alterations in cell lines *etc.* For every keyword, examples are also provided for instance upon clicking cell line BT474, all the assays done on BT474 cell line will be visible. In our web server, we have provided two modes of search:

- **Simple search**: This option provides general keyword search at top of all above mentioned fields. Here, a user can either select or provide partial text in search box for quering. This leads to all assay related information as selected for display.

- **Advanced search**: For extensive search with logical operators like AND, OR, exact or containing matching. For example, if the user is searching for all assays done on BT474 cell line and where cell line has been altered by inhibition of ADAM17, one can select these two options with AND logical operator.

The results in search options come in the form of a table, which gives assay details in initial columns as selected for display. In addition, for every search, the last nine columns show the genomic characteristics of that particular cell line as reported in CCLE database[14]. The genomic characteristics include expression of 22 important genes while last eight columns present mutation of eight important genes (as mentioned in method section).

Browse. We have provided several instructive and powerful browsing options, which provide an overall view on assay data. The unique feature of these browsing tables is that the user can sort and search the entries for every columns of result table. The browsing can be done based on following:

- **Browse on cell line**
  This facility bestows all the statistics of assay and genomic data keeping cell lines in mind. First eight columns present assay information pertaining to the number of assays done, drugs supplemented, alterations made in cell line *etc.* Second half of the table shows the genomic details of that cell line such as number of mutations reported in CCLE, genes mutated out of eight important genes, comparative expression of 22 important genes for every cell line and external link to CCLE database.
  Clicking on any of the link leads to a new table having details of that cell line, while keeping selected columns in mind. For example, BT474 has been studied in 43 different articles and 62 different drugs have been supplemented with it.

- **Browse on supplementary drugs/chemicals**
  This browsing menu presents details of 111 different drugs/chemicals. Here, the columns show the number of assays done with drugs and number of cell lines tested with this supplement.

- **Browse on alterations in cell lines**
  As many as 337 types of alterations reported in BCCs, have been mentioned in HerceptinR database. Each row mention the alteration, number of assays performed with that alteration and number of cell lines having this alteration.

- **Browse on PMID**
  The assay data acquired from 75 research articles can be browsed in this module. Here, for each research article (PMID), we have provided the numbers of assays, cell lines, drugs supplemented and alterations reported. The rows are also containing Title of article, links for PubMed and free full text, which can be downloaded in the form of PDF.

Cell line data. This section of database harbors genomic information of 51 BCCs. The genomic information is related to mutations, expression and CNV. In addition, drug sensitivity profile of various known drugs tested on these BCCs has also been included in this section. For better analysis and instructive study, we have developed following modules:

Mutation search. After comprehensive search of different BCCs for Herceptin response, a user can explore the mutational status of various key genes in that cell line. For this purpose, we have developed this module, where one can search queries for cell line, genes/proteins, cDNA mutation, protein mutation etc. In addition to this, the user can also select mutation by selecting criteria like protein family, domain or subcellular localization. For example, if the user needs to look at the mutation present in proteins of kinase superfamily of BT474 cell line, he can select cell line as 'BT474' and protein family as 'Protein kinase superfamily'. This selection will display the present mutation in 632 genes (as mentioned in method) of BT474 cell line.

Summary of cell line. This module gives details of genomic information of a cell line in its entirety. It includes a column for over-expressed genes, having 'expression value' greater than 12 (selected in the range of 0 to 15 in CCLE expression data). While the fourth column shows the under-expressed genes, having expression value 3 in above-mentioned scale. Two columns of all mutation and important gene mutation are same as cell line browsing. Seventh column of 'Drug Sensitivity plot' is one of the most important features, which imparts users a profile of already tested drugs on that cell lines and probable supplementary drugs for Herceptin administration. The drugs are plotted in their decreasing order of IC_{50}. Next to the plot, there are links to the CancerDR[15] database having all the drug sensitivity data.

Browse on multiple cell lines. As a very powerful and instructive tool, this browsing enables a user to compare cell lines on the basis of mutational, expression and CNV status of several cancer related genes. The browsing becomes more applicable by comparing Herceptin sensitive and resistant cell lines shown in green and red color respectively in selection table. This tool allows comparison of maximum five cell lines while selecting one feature at a time. The division of cell lines in to resistant and sensitive for genomic compare such as 'Browse on multiple cell lines', 'Relative GE/CNVs' and 'Compare genes', was done by taking those cell lines as resistant which are reported as resistant in 60% of its occurrence in the database and the rest were treated as sensitive.

Relative GE/CNVs (pair wise cell line compare of expression and CNV). Furthermore, to understand the mechanisms related to Herceptin resistance, we have also provided the pair wise comparison of cell lines, where a user can compare the expressional and CNV differences in two cell lines. With this tool, the user can look at the difference or ratio of CNV or expression of certain gene in selected cell lines. Based on difference or ratio, highly differing genes can be identified for CNV/expression. For example when we select cell lines C1 and C2 for pair wise comparison of expression, the result displays ratio of expression of genes like G1 by D1/D2 and subtraction of expression by D1–D2. The D1 and D2 values may be the numerical values of expression or CNV in cell line C1 and C2 respectively. In the past, in a comprehensive genomic study on breast cancer cell lines, authors[16] have ranked the expressing gene with in the cell line. We also adopted similar type of ranking strategy by providing the percentile ranking of expressing genes. This tool enables the user to identify genes among top expressing genes. Sorting the percentile ranking, difference and ratio with text searching makes the tool more revealing.

Compare genes. This tool provides a combined comparison of mutation, expression and CNV for any pair of cell line selected by the user.

Based on the findings of assay data, if the user needs to correlate the phenomenon of Herceptin resistance with the genomics of particular cell lines, he can select the cell lines and check the differences in all features such as mutation, expression and CNV. The expression and CNV of a gene can only be queried by comparing selected cell lines and percentile ranking given to them.

Alignment of mutants. To visualize and compare the mutations present in important mutant genes (see method section), we have provided a platform for multiple sequence alignment of all mutants of these genes with wild type sequence. For its visualization, we exploited the Jalview applet, which is a powerful and easy to use web based application. It enables the user to look at the aligned mutated regions on cDNA or protein sequences, which belong to a resistant or sensitive cell line.

Align my sequence. In continuity to alignment tool as described above, we have also provided the facility to align cDNA or protein sequence provided by the user. By selecting the mutant of choice or all mutants of a gene, a user can align and visualize the query sequence.

Discussion

Although Herceptin is effective in Her2 positive breast cancers, a considerable fraction of patients stop responding or lose clinical benefits by primary (*denovo*) or secondary (acquired) resistance respectively[4]. The mechanism and components rendering Herceptin resistance are ill defined so far[5]. Thus, exploration of potential biomarkers of Herceptin efficacy in HER2-positive breast cancer and evaluation of such markers to advice patient selection for therapy is of great value[12]. To understand the molecular mechanism and genomic factors contributing to the phenomenon of resistance against Herceptin, several studies have been carried out taking tissue sample or cell line based model systems[17,18]. Unfortunately, at present we do not have a single platform, where one can correlate experiments done with Herceptin with all the genomic factors such as mutation, expression and CNV, to elucidate the biomarkers for Herceptin resistance. Since cell lines are the established, homogeneous and well-studied models for cancer, a body of cell line data exists in the literature, which is related to Herceptin resistance and the supplementary drugs tested along with Herceptin. With this basis and aim to provide a platform, where comprehensive information is available on experiments aimed at checking efficacy of Herceptin alone and in combination with various drugs on various cancer cell lines. To have a general and gross overview we have provided search and browse options respectively. The significant feature, in both search and browse options, is the display of genomic feature (mutation and expression) of certain important genes, which are the characteristic of a particular cell line.

As another important information of assay data, we have provided alterations in cell lines, which can be understood by looking at the difference of two experiments. For example, BT474 cell line was assayed with Herceptin in two different experiments where in first it was treated with 'PERLD1 siRNA' while in second experiment 'ectopic expression of CYCLIN E' was done. Accordingly, we defined 'Silencing with PERLD1 siRNA' and 'ectopic expression of CYCLIN E' as alterations in cell lines. Such alterations become important while understanding the mechanism of resistance, locating new targets or assigning new supplementary drugs. The supplementary drugs included in our database are all those chemicals/drugs and supplements, which have been tested along with Herceptin in order to improve its efficacy in *in vitro* assays.

In the direction of supplementary drugs, our attempt to compile drug sensitivity profiles of different BCCs from CancerDR database, at the same platform, becomes very useful. As an extrapolation, the drugs having very high IC$_{50}$ reported could not be a good choice of supplementary drug with Herceptin. At the same time other drugs having very low IC$_{50}$ could be preferred drugs for supplement.

Furthermore, the searchable information on gene/protein and cell line related mutation enables a user to find out novel factors involved in resistance as many genes with mutations have been reported in cancer. Comparison of the genomic features of cell lines with respect to Herceptin resistance is the most instructive and comprehensive aspect of our database. Among various applications of this database, the collated information may pave the way for the development of Herceptin based personalized medicine for the breast cancer treatment. Currently, one of the limitations of our database is that HerceptinR does not provide any information on *in vivo* studies due to paucity of tissue data in the literature. But in the light of cell line data (Herceptin response information, genomics and drug profiles) this database provides valuable information about Herceptin resistance. In future, efforts will be made to provide similar information on cancer tissues samples.

Methods

Data construction. *Herceptin assay data.* The primary data on the assays performed with Herceptin were extracted from PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) using keyword "(herceptin OR trastuzumab) AND resistance AND breast cancer", with a period restriction of last than ten years. With these keywords and filters we obtained 277 free full text PubMed research articles. These articles were read carefully for assays performed with Herceptin, supplementary drugs, duration of treatment and other conditions, which differed from one experiment to other (Figure 1). In total, 31 cell lines were reported in final data from 75 research articles (listed on the web server) as shown in Figure 2. Since, we have to correlate the assay information with genomic information of the same entity and at the same time the genomics should be as homogeneous as possible, we focused on assays performed with cell lines only and experiments with tissue samples were not considered. The inhibition values were derived from different tables, graphs and text in results by either simply taking values or visual inspection of graphs. The response was categorized into resistance if either authors have mentioned it or there was less than 10% inhibition in assay. In other cases, the status is as such quoted as mentioned by the authors (sensitive or inhibition). Thus, our database includes three types of assay response: 1) Resistance, 2) Sensitive and 3) Inhibition.

Cell line (genomic) data. As a second important dimension of data, we acquired genomic data for 51 BCCs which were available in CCLE (Cancer Cell Line Encyclopedia)[14] and CancerDR[15] databases, as shown in Figure 2 & Figure 3. The study of CCLE includes high throughput sequencing of 1650 cancer genes of 904 cancer cell lines and expression analysis of more than 16582 genes for 550 cancer cell lines. Similarly CNV for 16582 genes were obtained for 998 cancer cell lines in CCLE study. In addition to this drug sensitivity data for more than 500 cancer cell lines was obtained from CancerDR. As a subset of whole CCLE dataset and CancerDR data, our genomic data only included breast cancer cell lines. Finally, our genomic data comprised of following types of data:

- **Mutation data**: The CCLE (http://www.broadinstitute.org/ccle/) maintain the mutation data that include the mutational profile of important cancer related genes. We procured the data of 632 genes, which were reported in any of the BCCs in CCLE. The selection of genes in CCLE was based on: 1) occurrence in at least 4 instances in research article or Cancer Gene Sensus, 2) frequency of occurrence i.e in 441 tumors in SEER database, 3) functionality e.g. oncogenes, component of cancer pathway or tumor suppressor genes etc. The sequencing was reported to be done by Hybrid capture sequencing.

- **Gene expression**: Similar to mutation data, the RMA-normalized expression data of 16582 genes is available in CCLE where expression ranges between levels 0 to 15. For our study, we extracted expression of these genes for 51 BCCs. The authors of CCLE database[14] obtained the mRNA expression data using Affymetrix Human Genome U133 Plus 2.0 arrays as per the manufacturer's instructions. The background correction was accomplished by RMA (Robust Multichip Average) and quantile normalization. Since the expression value is given in log$_2$ of expression, the values ranged from lowest expression value of 0 to maximum of 15 where change of expression from say 7 to 8 refers to increase in expression by two folds. Such variations of expression of certain gene along different cell line can be used for investigation of factors responsible for Herceptin resistance. To provide a gross view on expression of genes with in cell line as done in other similar studies in the past[16], we have calculated percentile rank for all the genes of expression data The ranking provides the place of a gene among top or bottom expressing genes. The percentile rank was calculated by following formula:

$$Pg1 = \frac{E+B}{N} * 100$$

Where P$_{g1}$ is percentile score of gene *g1*, B is number of genes having expression less than gene *g1*, *E is* number of genes having expression equal to gene *g1*, *N* is total number of genes within the given cell line.
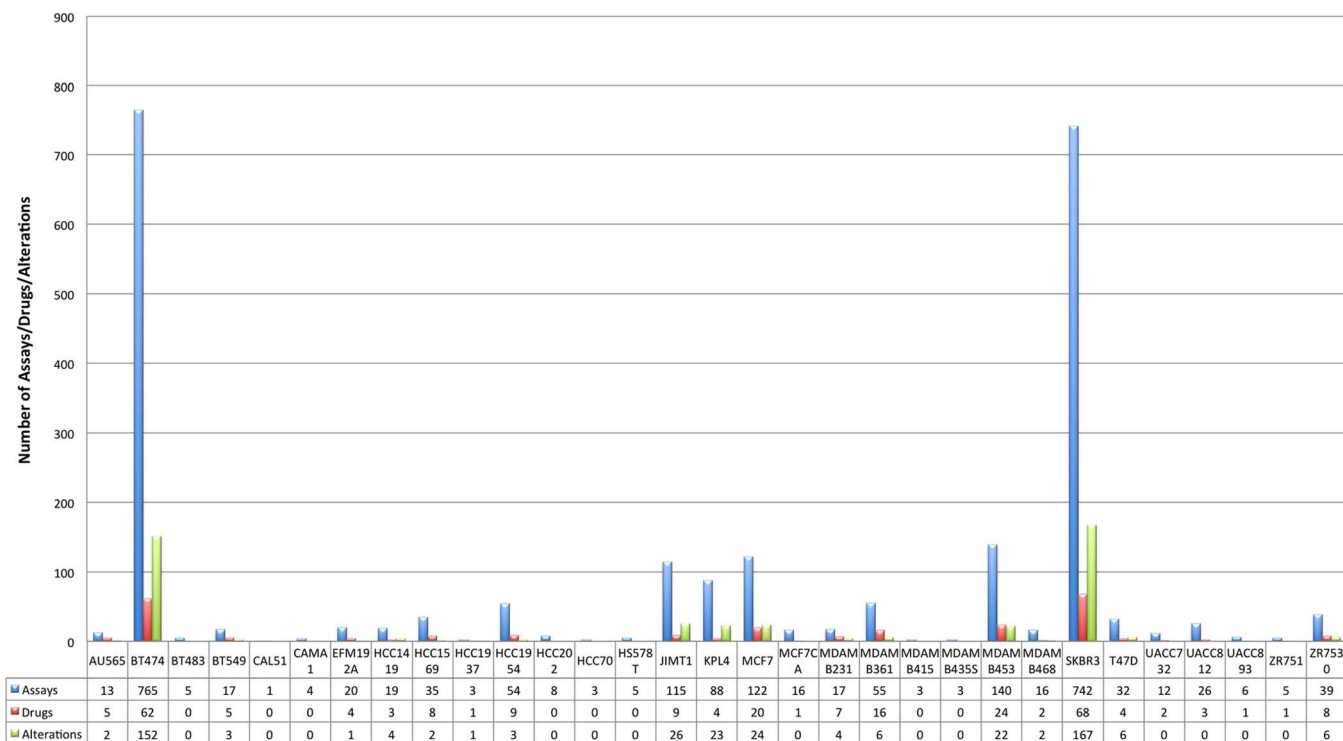
Figure 1 | Schematic diagram showing distribution of assays, supplementary drugs and alterations over different breast cancer cell lines.
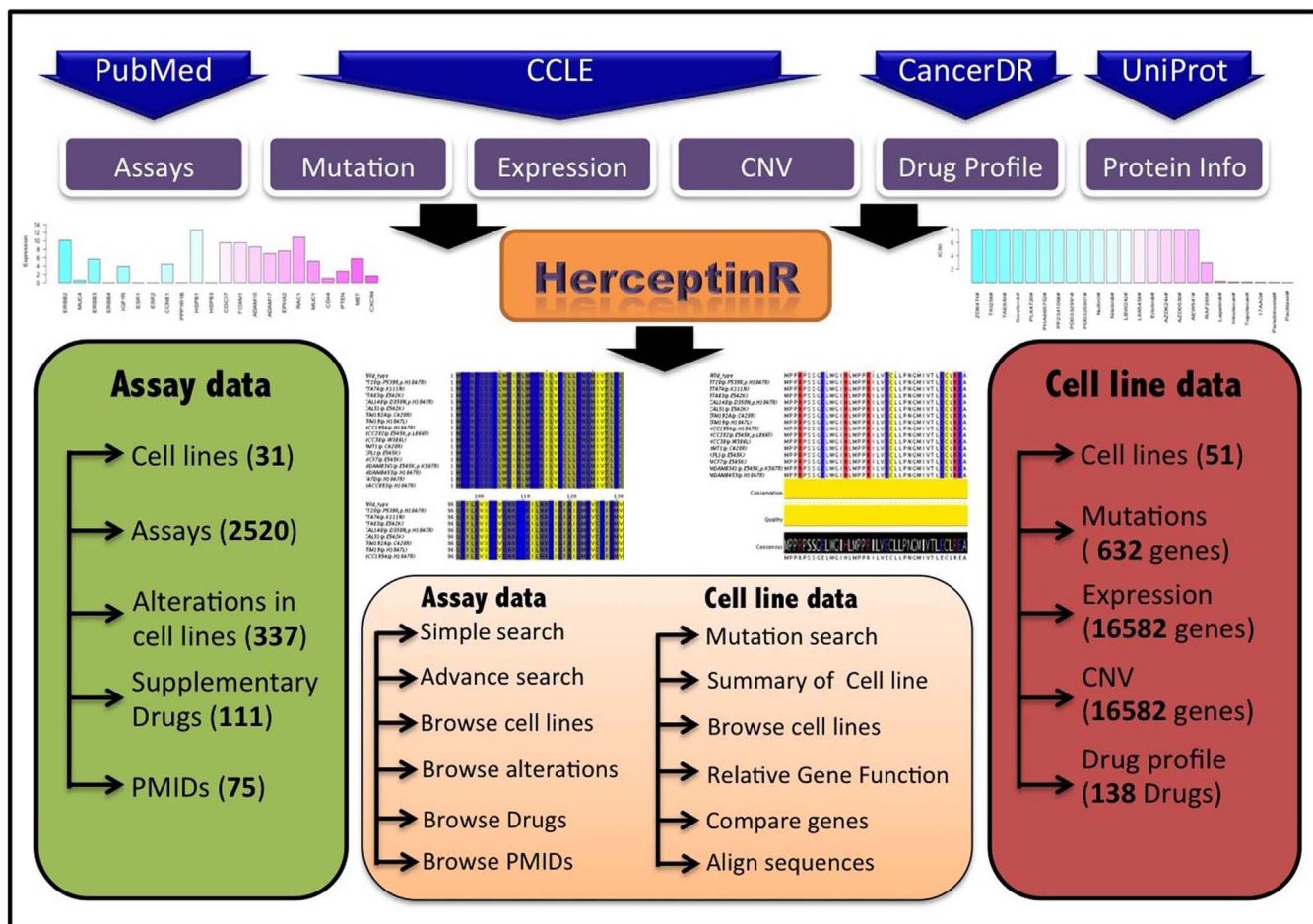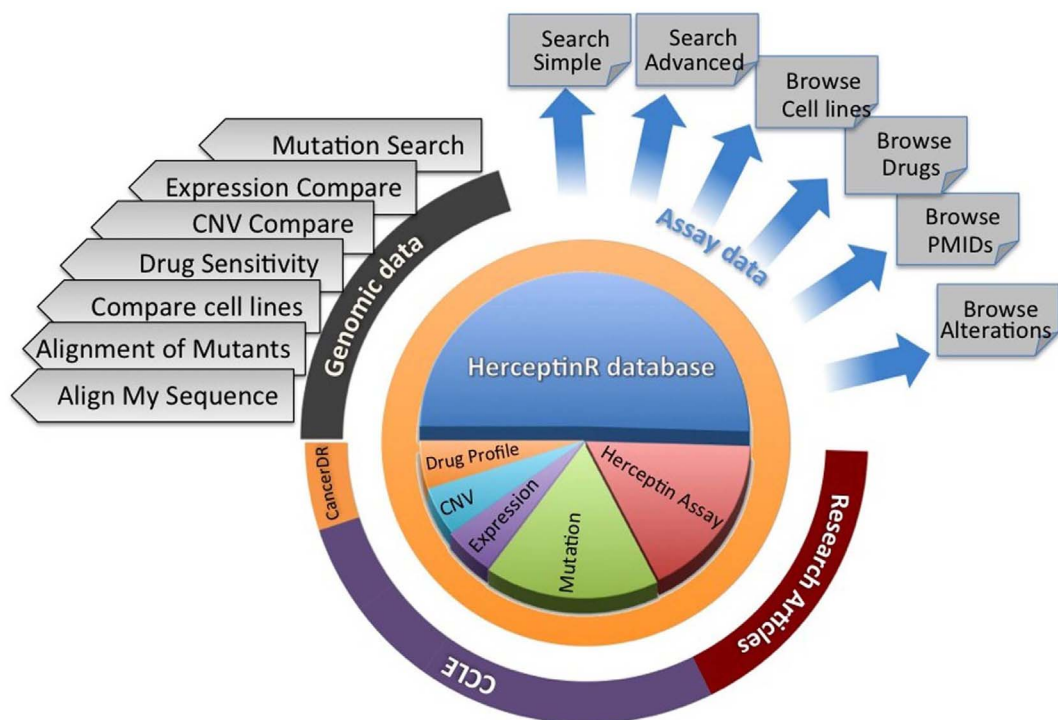
	AU565	BT474	BT483	BT549	CAL51	CAMA1	EFM192A	HCC1419	HCC1569	HCC1937	HCC1954	HCC202	HCC70	HS578T	JIMT1	KPL4	MCF7	MCF7CA	MDAMB231	MDAMB361	MDAMB415	MDAMB435S	MDAMB453	MDAMB468	SKBR3	T47D	UACC732	UACC812	UACC893	ZR751	ZR7530
Assays	13	765	5	17	1	4	20	19	35	3	54	8	3	5	115	88	122	16	17	55	3	3	140	16	742	32	12	26	6	5	39
Drugs	5	62	0	5	0	0	4	3	8	1	9	0	0	0	9	4	20	1	7	16	0	0	24	2	68	4	2	3	1	1	8
Alterations	2	152	0	3	0	0	1	4	2	1	3	0	0	0	26	23	24	0	4	6	0	0	22	2	167	6	0	0	0	0	6

Figure 2 | Schematic illustartion of architecture of HerceptinR.

Figure 3 | Diagram showing various applications of HerceptinR.

Since the normalized expression values for every gene, ranged from 0–15, we tried to categorize the expressing genes in two classes: 1) Over-expressed (having expression value greater than 12); 2) Under-expressed (having expression value lower than 3). We used these classes in 'Summary of cell lines module'.

- **Copy number variation**: We obtained copy number variation (CNV) data for 16582 genes of BCCs from CCLE. According to the authors[14], the raw Affymetrix CEL files were converted to a single value for each probe set representing a SNP allele or a copy number probe. Copy numbers can be understood as $\log_2$ of ratio of copies of gene in normal vs. cancer. Thus the positive and negative values of CNV signifies the increase and decrease in copies of gene respectively. To further provide cell line wise significant genes based on CNV, we calculated percentiles of every gene for BCCs of compare modules in similar manner as expression data. The percentile ranking fascilitate the user to explore genes amplified and present in top genes say top 25% of genes, as done in other similar studies in the past[16].

- **Drug sensitivity data**: Considering 138 drugs, reported in CancerDR database[15], as probable and novel supplementary drugs in Herceptin treatment, we extracted the drug sensitivity data for all the 31 BCCs mentioned in Herceptin assay data section. This data includes the pharmacological profiling [IC_{50}] of 138 important anticancer drugs on BCCs.

- **Protein information**: We picked up protein description, protein family, domain and subcellular localization information of all mutated proteins from UniProt database[19]. These descriptions of proteins belong to normal and healthy cells.

Important genes involved in resistance. As per different reports, there are number of important genes, which have been hypothesized to be involved in resistance against Herceptin. The contribution of a gene may be via two different ways:

- **Gene expression**: Several genes like ERBB2[20], MUC4[21], ERBB3[22,23], ERBB4[24], IGF1R[25], ESR1[26], ESR2[26], CCNE1[27], PPP1R1B[28], HSPB1[29], HSPB3[30], CDC37[31], FOXM1[32], ADAM10[33], ADAM17[34], EPHA2[35], RAC1[36,37], MUC1[38], CD44[39], PTEN[6,40], MET[41], CXCR4[42], reported, to be involved in resistance by their altered expression. We have taken the expression values of these 22 genes from CCLE database for different BCCs and presented in the form of expression plot as the expression characteristic of that cell line.

- **Mutation data**: It has been reported in past that mutation of certain genes play vital role in Herceptin resistance. Thus, we incorporate mutational status of 8 such genes, namely- PIK3CA[43], PTEN, RB1[44], TP53[45], BRCA1[46], BRCA2[46], MAP2K4[47], and MAP3K1[6], for each of 51 BCCs from CCLE database. Mutational status of these genes were taken as characteristic of every cell line and presented with cell line in many result tables.

Mutants and alignment. We picked up eight important genes (as mentioned above in important gene-mutation) from the mutational data, mapped all the mutations present in that gene, in that particular cell line and aligned all those mutants [cDNA

and proteins] with Clustal-W[48]. The mutational information of eight genes is displayed with every assay entry as a characteristic for that particular cell line. In the alignment section, we have made use of Jalview applet (http://www.jalview.org/) for visualization of these mutants[49].

Database architecture and web interface. HerceptinR is developed with the help of Apache HTTP server 2.2 with MySQL 5.1.47 at the back end and the PHP 5.2.9, HTML and JavaScript at the front end. Being open-source and platform independent software, Apache, MySQL, and PHP are preferred. Expression and drug sensitivity plots were made with R 2.15.1 package (www.r-project.org/).

1. Rubin, I. & Yarden, Y. The basic biology of HER2. *Ann. Oncol.* **12 Suppl 1**, S3–8 (2001).
2. Morrow, P. K. H., Zambrana, F. & Esteva, F. J. Recent advances in systemic therapy: Advances in systemic therapy for HER2-positive metastatic breast cancer. *Breast Cancer Res.* **11**, 207 (2009).
3. Vu, T. & Claret, F. X. Trastuzumab: updated mechanisms of action and resistance in breast cancer. *Front. Oncol.* **2**, 62 (2012).
4. Pohlmann, P. R., Mayer, I. A. & Mernaugh, R. Resistance to Trastuzumab in Breast Cancer. *Clin. Cancer Res.* **15**, 7479–7491 (2009).
5. De, P., Hasmann, M. & Leyland-Jones, B. Molecular determinants of trastuzumab efficacy: What is their clinical relevance? *Cancer Treat. Rev.* **39**, 925–34 (2013).
6. Duman, B. B., Sahin, B., Acikalin, A., Ergin, M. & Zorludemir, S. PTEN, Akt, MAPK, p53 and p95 expression to predict trastuzumab resistance in HER2 positive breast cancer. *J. BUON.* **18**, 44–50.
7. Kataoka, Y. *et al.* Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines. *Ann. Oncol.* **21**, 255–62 (2010).
8. Park, B. H. & Davidson, N. E. PI3 kinase activation and response to Trastuzumab Therapy: what's neu with herceptin resistance? *Cancer Cell* **12**, 297–9 (2007).
9. Hicks, D. G. & Kulkarni, S. Trastuzumab as adjuvant therapy for early breast cancer: the importance of accurate human epidermal growth factor receptor 2 testing. *Arch. Pathol. Lab. Med.* **132**, 1008–15 (2008).
10. Zografos, G. C. & Roukos, D. H. Innovative biomarker development for personalized medicine in breast cancer care. *Biomark. Med.* **5**, 73–8 (2011).
11. Kelloff, G. J. & Sigman, C. C. Cancer biomarkers: selecting the right drug for the right patient. *Nat. Rev. Drug Discov.* **11**, 201–14 (2012).
12. Cayre, A., Mishellany, F., Lagarde, N. & Penault-Llorca, F. Comparison of different commercial kits for HER2 testing in breast cancer: looking for the accurate cutoff for amplification. *Breast Cancer Res.* **9**, R64 (2007).
13. Lourenço, H. M., Pereira, T. P., Fonseca, R. R. & Cardoso, P. M. HER2/neu detection by immunohistochemistry: optimization of in-house protocols. *Appl. Immunohistochem. Mol. Morphol.* **17**, 151–7 (2009).
14. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–7 (2012).

15. Kumar, R. *et al.* CancerDR: cancer drug resistance database. *Sci. Rep.* **3**, 1445 (2013).
16. Zaman, N. *et al.* Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep.* **5**, 216–23 (2013).
17. Gajria, D. & Chandarlapaty, S. HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Rev. Anticancer Ther.* **11**, 263–75 (2011).
18. Nahta, R., Yu, D., Hung, M.-C., Hortobagyi, G. N. & Esteva, F. J. Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer. *Nat. Clin. Pract. Oncol.* **3**, 269–80 (2006).
19. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol. Biol.* **406**, 89–112 (2007).
20. Mitra, D. *et al.* An oncogenic isoform of HER2 associated with locally disseminated breast cancer and trastuzumab resistance. *Mol. Cancer Ther.* **8**, 2152–62 (2009).
21. Nagy, P. *et al.* Decreased accessibility and lack of activation of ErbB2 in JIMT-1, a herceptin-resistant, MUC4-expressing breast cancer cell line. *Cancer Res.* **65**, 473–82 (2005).
22. Huang, X. *et al.* Heterotrimerization of the growth factor receptors erbB2, erbB3, and insulin-like growth factor-i receptor in breast cancer cells resistant to herceptin. *Cancer Res.* **70**, 1204–14 (2010).
23. Wang, S. E. *et al.* Transforming growth factor beta engages TACE and ErbB3 to activate phosphatidylinositol-3 kinase/Akt in ErbB2-overexpressing breast cancer and desensitizes cells to trastuzumab. *Mol. Cell. Biol.* **28**, 5605–20 (2008).
24. Sassen, A. *et al.* Presence of HER4 associates with increased sensitivity to Herceptin in patients with metastatic breast cancer. *Breast Cancer Res.* **11**, R50 (2009).
25. Lu, Y., Zi, X. & Pollak, M. Molecular mechanisms underlying IGF-I-induced attenuation of the growth-inhibitory activity of trastuzumab (Herceptin) on SKBR3 breast cancer cells. *Int. J. Cancer* **108**, 334–41 (2004).
26. Hevir, N., Trošt, N., Debeljak, N. & Rižner, T. L. Expression of estrogen and progesterone receptors and estrogen metabolizing enzymes in different breast cancer cell lines. *Chem. Biol. Interact.* **191**, 206–16 (2011).
27. Scaltriti, M. *et al.* Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2+ breast cancer patients. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3761–6 (2011).
28. The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells. *Mol. Oncol.* **7**, 392–401 (2013).
29. Kang, S. H. *et al.* Upregulated HSP27 in human breast cancer cells reduces Herceptin susceptibility by increasing Her2 protein stability. *BMC Cancer* **8**, 286 (2008).
30. Vos, M. J., Kanon, B. & Kampinga, H. H. HSPB7 is a SC35 speckle resident small heat shock protein. *Biochim. Biophys. Acta* **1793**, 1343–53 (2009).
31. Stepanova, L., Leng, X., Parker, S. B. & Harper, J. W. Mammalian p50Cdc37 is a protein kinase-targeting subunit of Hsp90 that binds and stabilizes Cdk4. *Genes Dev.* **10**, 1491–502 (1996).
32. Carr, J. R., Park, H. J., Wang, Z., Kiefer, M. M. & Raychaudhuri, P. FoxM1 mediates resistance to herceptin and paclitaxel. *Cancer Res.* **70**, 5054–63 (2010).
33. Liu, P. C. C. *et al.* Identification of ADAM10 as a major source of HER2 ectodomain sheddase activity in HER2 overexpressing breast cancer cells. *Cancer Biol. Ther.* **5**, 657–64 (2006).
34. Gijsen, M. *et al.* HER2 phosphorylation is maintained by a PKB negative feedback loop in response to anti-HER2 herceptin in breast cancer. *PLoS Biol.* **8**, e1000563 (2010).
35. Zhuang, G. *et al.* Elevation of receptor tyrosine kinase EphA2 mediates resistance to trastuzumab therapy. *Cancer Res.* **70**, 299–308 (2010).
36. Inactivation of Rac1 reduces Trastuzumab resistance in PTEN deficient and insulin-like growth factor I receptor overexpressing human breast cancer SKBR3 cells. *Cancer Lett.* **313**, 54–63 (2011).
37. Dokmanovic, M., Hirsch, D. S., Shen, Y. & Wu, W. J. Rac1 contributes to trastuzumab resistance of breast cancer cells: Rac1 as a potential therapeutic target for the treatment of trastuzumab-resistant breast cancer. *Mol. Cancer Ther.* **8**, 1557–69 (2009).
38. Fessler, S. P., Wotkowicz, M. T., Mahanta, S. K. & Bamdad, C. MUC1* is a determinant of trastuzumab (Herceptin) resistance in breast cancer cells. *Breast Cancer Res. Treat.* **118**, 113–24 (2009).
39. Oliveras-Ferraros, C. *et al.* Dynamic emergence of the mesenchymal CD44(pos)CD24(neg/low) phenotype in HER2-gene amplified breast cancer cells with de novo resistance to trastuzumab (Herceptin). *Biochem. Biophys. Res. Commun.* **397**, 27–33 (2010).
40. Faratian, D. *et al.* Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab. *Cancer Res.* **69**, 6713–20 (2009).
41. Shattuck, D. L., Miller, J. K., Carraway, K. L. & Sweeney, C. Met receptor contributes to trastuzumab resistance of Her2-overexpressing breast cancer cells. *Cancer Res.* **68**, 1471–7 (2008).
42. Gros, S. J. *et al.* Involvement of CXCR4 chemokine receptor in metastatic HER2-positive esophageal cancer. *PLoS One* **7**, e47287 (2012).
43. The predictive role of phosphatase and tensin homolog (PTEN) loss, phosphoinositol-3 (PI3) kinase (PIK3CA) mutation, and PI3K pathway activation in sensitivity to trastuzumab in HER2-positive breast cancer: a meta-analysis. *Curr. Med. Res. Opin.* **29**, 633–42 (2013).
44. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108–12 (2013).
45. Relationship between NF-κB, ER, PR, Her2/neu, Ki67, p53 expression in human breast cancer. *Exp. Oncol.* **34**, 358–63 (2012).
46. Fasano, J. & Muggia, F. Breast cancer arising in a BRCA-mutated background: therapeutic implications from an animal model and drug development. *Ann. Oncol.* **20**, 609–14 (2009).
47. Su, G. H., Song, J. J., Repasky, E. A., Schutte, M. & Kern, S. E. Mutation rate of MAP2K4/MKK4 in breast carcinoma. *Hum. Mutat.* **19**, 81 (2002).
48. Higgins, D. G., Thompson, J. D. & Gibson, T. J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402 (1996).
49. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–91 (2009).

Acknowledgments

Author contributions

S.A. and S.G. collected and organized the data. S.G. and R.K. developed web interface. S.G. and S.A. contributed in manuscript writing. G.P.S.R. and G.C.V. conceived the idea and coordinated the project.

Additional information