Integrated Analysis of Copy Number and Gene Expression

Nexus Copy Number provides user-friendly interface and functionalities to integrate copy number analysis with gene expression results for the purpose of identifying the "Genomic Hotspots" containing both copy number aberrations and changes in gene expression

A WHITE PAPER FROM

BioDiscovery, Inc.

September 2009

Introduction

DNA amplification and up-regulation of critical oncogenes are important mechanisms for cancer proliferation. In breast cancer for example, ERBB2, MYC, and CCND1 as well as other genes and regions are known to be amplified in a higher percentage of the tumor samples and are usually associated with late stages of the disease. The gene expression profiles of these genes and their relationships with breast cancer have also been well studied. However, the systematic investigation of both DNA aberrations and gene expression has just started in recent years, and the complicated data analysis has made this process a daunting task for scientists.

Nexus Copy Number is not only a software tool designed for scientists to analyze copy number variations, but also a tool providing an easy and straightforward way to integrate gene expression results (differentially up- and down-regulated gene lists). By combining gene expression changes with copy number gains or losses, "Genomic Hotspots" that show significant correlations between gene expression and CNV can be identified.

The Breast Cancer Data Set

Agilent 244K Feature Extraction data files from 58 FFPE breast cancer samples were kindly provided by Dr. Jeffrey Gregg at UC Davis. Raw data in the form of Affymetrix Mapping 500K CEL files for a set of 49 breast cancer and 2 Normal fresh frozen samples were downloaded from the paper *High-resolution genomic and expression analyses of copy number alterations in breast tumors* (Haverty, 2008). These two data sets were loaded into Nexus Copy Number and the copy number variation profiles for all 109 samples are displayed in Figure 1.

Gene expression data was downloaded from a public database associated with the paper *Concordance among gene-expression-based predictors for breast cancer* (Fan, 2006). It has 70 ER negative and 225 ER positive samples and the data has been analyzed in BioDiscovery's Nexus Expression software to identify a set of 1703 significantly differentially expressed genes between ER positive and negative samples (False Discovery Rate corrected q-bound<0.01). A hierarchical clustering of the samples and genes is shown in Figure 2.

Fig. 1. A total of 109 breast cancer samples (51 fresh frozen samples on Affymetrix 500K mapping arrays and 58 FFPE samples on Agilent 244K arrays) were imported and processed in Nexus Copy Number. The frequency plots for all samples (top) and the two array types (bottom) are displayed with copy number gains in green and losses in red.

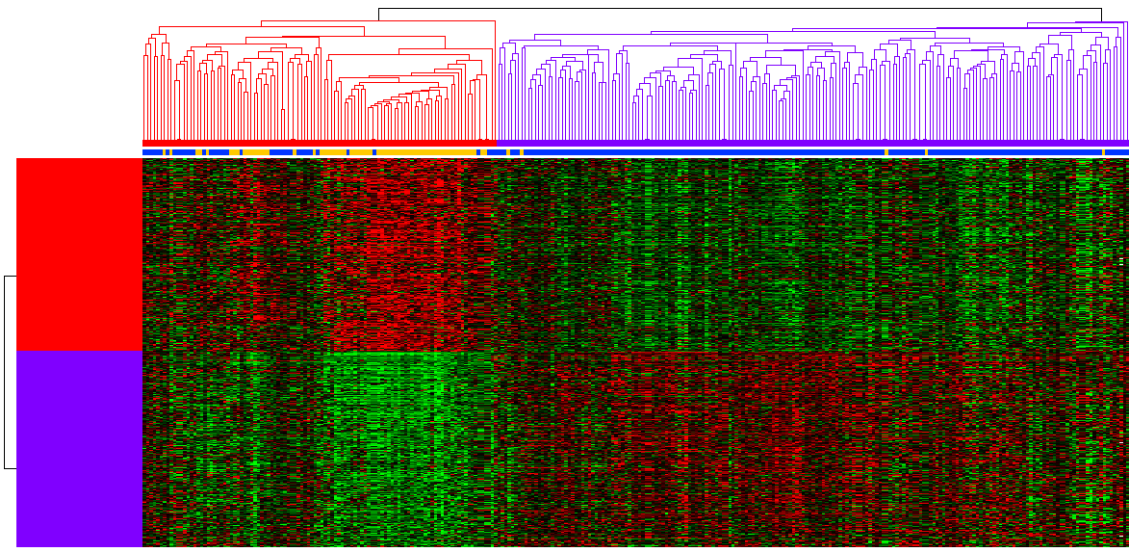


Fig. 2. Heat map of 1073 genes found to be differentially expressed in ER positive vs. negative samples. Hierarchical clustering shows that the samples mostly can be effectively separated into correct classes.

Integrating Gene Expression Results

The differentially regulated genes shown in Figure 2 were exported and saved in a tab-delimited text file. This file can be imported into Nexus Copy Number easily by clicking on the **Add** button in the **Expression** sub-tab Under **External Data** (Figure 3). In fact, gene lists generated using any gene expression software tool can be imported as long as they are in the correct text file format, consisting of columns **Gene Symbol, Regulation, Log-ratio,** and **p-value**. As shown in Figure 3, the gene list from ER Positive vs. Negative as well as other gene lists has been imported into Nexus Copy Number.

The gene expression profiles represented in these gene lists can be viewed along with the copy number frequency plot in Nexus Copy Number (Figure 4).

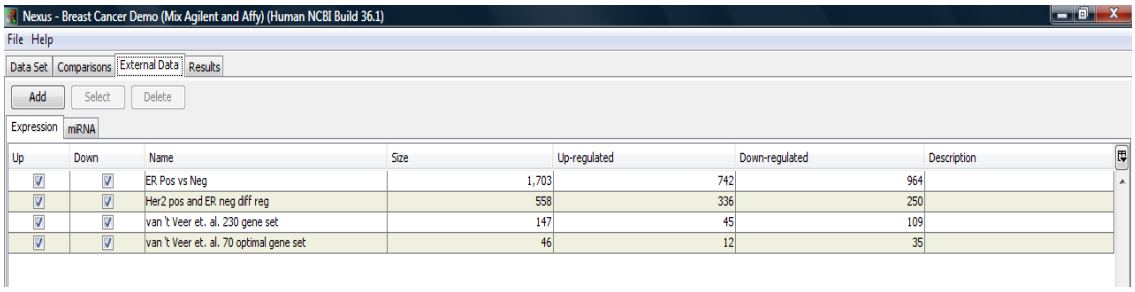


Fig. 3. Differentially regulated genes, generated by Nexus Expression, by any other gene expression software tool, or even downloaded from public databases, are easily imported into Nexus Copy Number. The 1703 Genes from ER Positive vs. Negative is shown as the first item on the list.

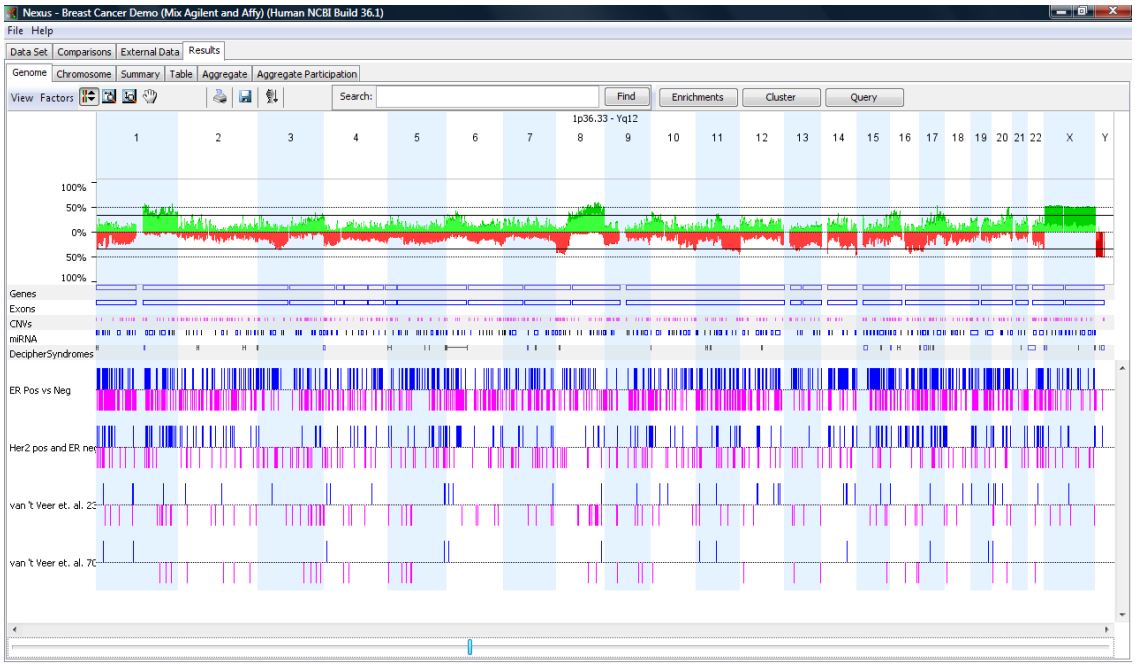
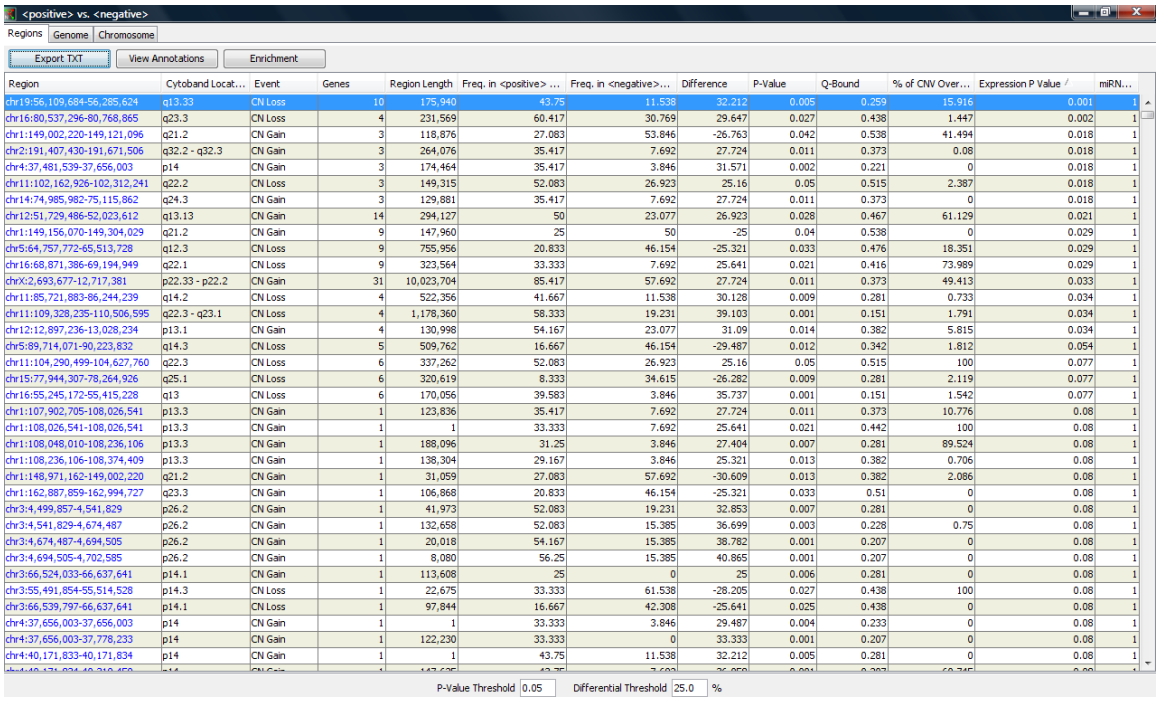


Fig. 4. Differentially regulated genes are visualized along with the CNV frequency plot for all the samples. Nexus Copy Number provides an easy interface to display both copy number and gene expression results and allows users to drill down to interesting chromosomal regions.

Locating “Genomic Hotspots”

The so-called “Genomic Hotspots” are regions on the chromosome that show significant correlation between copy number and gene expression changes. Nexus Copy Number can easily locate significantly different CNV regions between any two groups with its **Comparison** feature. With gene expression results imported, an **Expression P Value** is calculated for each significant CNV region that contains a significant group of genes with differential regulation. In the example shown in Figure 5, the region chr19:56,109,684-56,285,624 shows a copy number loss between ER Positive (43.75% of the samples showing a loss) and ER Negative (11.53% of the samples showing a loss), meanwhile an **Expression P Value** of 0.001 was calculated for this region. Five of the 10 genes in this region display significant down regulations (Figure 6). It is interesting to note that the genes are of the human kallikrein (KLK) gene family which have been found to be down regulated in ER positive (conversely, up regulated in ER negative) patients and have been used as prognostic biomarkers in breast cancer studies.



Region	Cytoband Locat...	Event	Genes	Region Length	Freq. in <positive> ...	Freq. in <negative>...	Difference	P-Value	Q-Bound	% of CNV Over...	Expression P Value	miRN...
chr19:56,109,684-56,285,624	q13.33	CN Loss	10	175,940	43.75	11.538	32.212	0.005	0.259	15.916	0.001	1
chr16:80,537,296-80,768,865	q23.3	CN Loss	4	231,569	60.417	30.769	29.647	0.027	0.438	1.447	0.002	1
chr1:149,002,220-149,121,096	q21.2	CN Gain	3	118,876	27.083	53.846	-26.763	0.042	0.538	41.494	0.018	1
chr2:191,407,430-191,671,506	q32.2 - q32.3	CN Gain	3	264,076	35.417	7.692	27.724	0.011	0.373	0.08	0.018	1
chr4:37,481,539-37,656,003	p14	CN Gain	3	174,464	35.417	3.846	31.571	0.002	0.221	0	0.018	1
chr11:102,162,926-102,312,241	q22.2	CN Loss	3	149,315	52.083	26.923	25.16	0.05	0.515	2.387	0.018	1
chr14:74,985,982-75,115,862	q24.3	CN Gain	3	129,881	35.417	7.692	27.724	0.011	0.373	0	0.018	1
chr12:51,729,486-52,023,612	q13.13	CN Gain	14	294,127	50	23.077	26.923	0.028	0.467	61.129	0.021	1
chr1:149,156,070-149,304,029	q21.2	CN Gain	9	147,960	25	50	-25	0.04	0.538	0	0.029	1
chr5:64,757,772-65,513,728	q12.3	CN Loss	9	755,956	20.833	46.154	-25.321	0.033	0.476	18.351	0.029	1
chr16:68,871,386-69,194,949	q22.1	CN Loss	9	323,564	33.333	7.692	25.641	0.021	0.416	73.989	0.029	1
chrX:2,693,677-12,717,381	p22.33 - p22.2	CN Gain	31	10,023,704	85.417	57.692	27.724	0.011	0.373	49.413	0.033	1
chr11:85,721,883-86,244,239	q14.2	CN Loss	4	522,356	41.667	11.538	30.128	0.009	0.281	0.733	0.034	1
chr11:109,328,235-110,506,595	q22.3 - q23.1	CN Loss	4	1,178,360	58.333	19.231	39.103	0.001	0.151	1.791	0.034	1
chr12:12,897,236-13,028,234	p13.1	CN Gain	4	130,998	54.167	23.077	31.09	0.014	0.382	5.815	0.034	1
chr5:89,714,071-90,223,832	q14.3	CN Loss	5	509,762	16.667	46.154	-29.487	0.012	0.342	1.812	0.054	1
chr11:104,290,499-104,627,760	q22.3	CN Loss	6	337,262	52.083	26.923	25.16	0.05	0.515	100	0.077	1
chr15:77,944,307-78,264,926	q25.1	CN Loss	6	320,619	8.333	34.615	-26.282	0.009	0.281	2.119	0.077	1
chr16:55,245,172-55,415,228	q13	CN Loss	6	170,056	39.583	3.846	35.737	0.001	0.151	1.542	0.077	1
chr1:107,902,705-108,026,541	p13.3	CN Gain	1	123,836	35.417	7.692	27.724	0.011	0.373	10.776	0.08	1
chr1:108,026,541-108,026,541	p13.3	CN Gain	1	1	33.333	7.692	25.641	0.021	0.442	100	0.08	1
chr1:108,048,010-108,236,106	p13.3	CN Gain	1	188,096	31.25	3.846	27.404	0.007	0.281	89.524	0.08	1
chr1:108,236,106-108,374,409	p13.3	CN Gain	1	138,304	29.167	3.846	25.321	0.013	0.382	0.706	0.08	1
chr1:148,971,162-149,002,220	q21.2	CN Gain	1	31,059	27.083	57.692	-30.609	0.013	0.382	2.086	0.08	1
chr1:162,887,859-162,994,727	q23.3	CN Gain	1	106,868	20.833	46.154	-25.321	0.033	0.51	0	0.08	1
chr3:4,499,857-4,541,829	p26.2	CN Gain	1	41,973	52.083	19.231	32.853	0.007	0.281	0	0.08	1
chr3:4,541,829-4,674,487	p26.2	CN Gain	1	132,658	52.083	15.385	36.699	0.003	0.228	0.75	0.08	1
chr3:4,674,487-4,694,505	p26.2	CN Gain	1	20,018	54.167	15.385	38.782	0.001	0.207	0	0.08	1
chr3:4,694,505-4,702,585	p26.2	CN Gain	1	8,080	56.25	15.385	40.865	0.001	0.207	0	0.08	1
chr3:66,524,033-66,637,641	p14.1	CN Gain	1	113,608	25	0	25	0.006	0.281	0	0.08	1
chr3:55,491,854-55,514,528	p14.3	CN Loss	1	22,675	33.333	61.538	-28.205	0.027	0.438	100	0.08	1
chr3:66,539,797-66,637,641	p14.1	CN Loss	1	97,844	16.667	42.308	-25.641	0.025	0.438	0	0.08	1
chr4:37,656,003-37,656,003	p14	CN Gain	1	1	33.333	3.846	29.487	0.004	0.233	0	0.08	1
chr4:37,656,003-37,778,233	p14	CN Gain	1	122,230	33.333	0	33.333	0.001	0.207	0	0.08	1
chr4:40,171,833-40,171,834	p14	CN Gain	1	1	43.75	11.538	32.212	0.005	0.281	0	0.08	1

Fig. 5. Significantly different CNV regions from the comparison of ER Positive vs. Negative. An Expression P value for each region was calculated based on the number of differentially regulated genes in the region. Sorting the Expression P Value column easily locates the CNV regions that are highly correlated with gene expression, the so-called “Genomic Hotspots”.

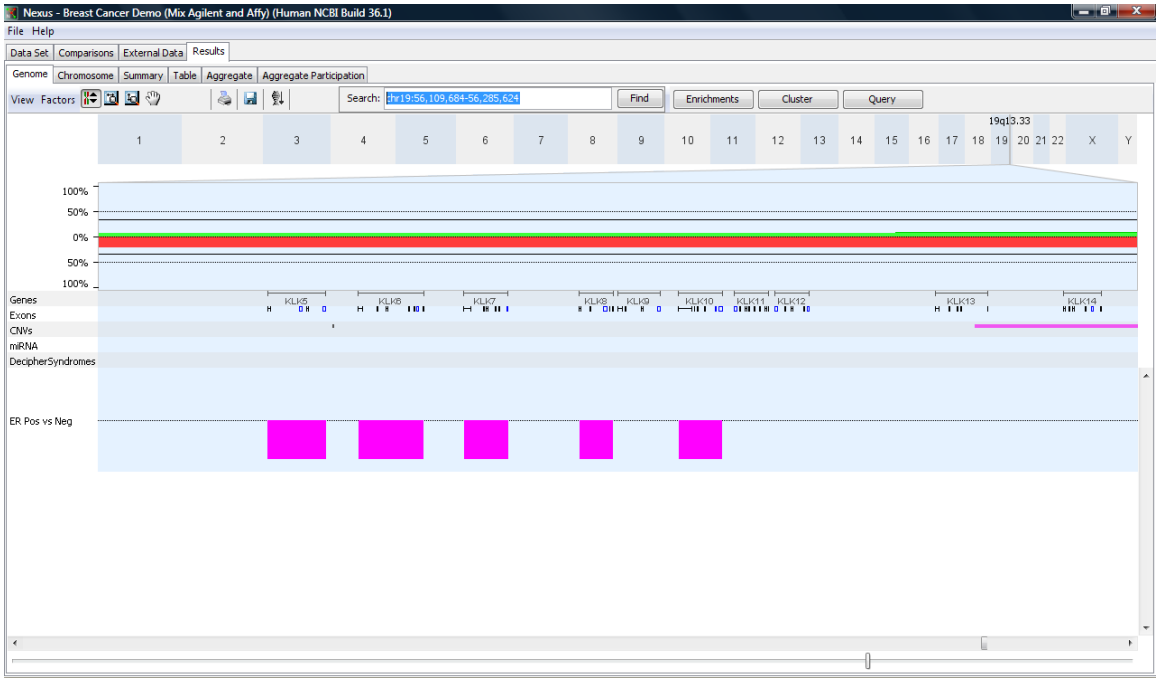


Fig. 6. Five of the 10 human kallikrein (KLK) genes show down regulations in one of the ER Positive vs. Negative CNV regions (chr19:56,109,684-56,285,624). This region is labeled as one of the "Genomic Hotspots". The human kallikrein (KLK) gene family has been found to be an important prognostic biomarker for breast cancer.

Conclusion

In addition to many advanced copy number analysis features, Nexus Copy Number provides scientists an easy and straightforward way to import gene expression results, integrate with the analysis of copy number variations, and locate CNV regions that significantly correlate with differentially regulated genes - "Genomic Hotspots". Since Integrated Genomics (which encompass studying copy number variations and gene expression together) has become the trend in genomic research, great value lies in Nexus Copy Number with which scientists can easily search for potential biomarkers themselves.

Additional Information

For more information about Nexus Copy Number, to request an online demo, or to download a trial of the software, please visit www.biodiscovery.com/index/nexus. For a complete list of our microarray software solutions, please visit us at www.biodiscovery.com or contact us at the address below.

BioDiscovery, Inc.
2221 Rosecrans Ave.
Suite 227
El Segundo, CA 90245
Tel: 310-414-8100
FAX: 310-414-8111
Email: info@biodiscovery.com
Website: www.biodiscovery.com

BioDiscovery, Inc.
2221 Rosecrans Ave. Suite 227
El Segundo, CA 90245
USA

Tel: +1 310-414-8100
FAX: +1 310-414-8111
Email: info@biodiscovery.com
Web: www.biodiscovery.com