

Cancer Survival and Copy Number Variations

The Predicative Power feature in Nexus Copy Number enables researchers to easily correlate the survival time of cancer patients to the significant CNV regions on their chromosomes

A WHITE PAPER FROM

BioDiscovery, Inc.

September 2009

Introduction

Cancer survival studies are useful in assessing the effectiveness of a treatment as well as in identifying potential prognostic markers and therapeutic targets. In the past few years, the association between copy number events in certain regions of the chromosomes and cancer survival time has been investigated by researchers in the aCGH field. The data analysis, though, poses a great challenge for researchers. For research groups without bioinformatics support, it may take a long time and much effort to obtain results from this type of analysis. Even for large institutions with bioinformatics groups, it still may require constant interaction between researchers and bioinformaticians.

The paper *Genome-Wide Copy Number Analysis in Esophageal Adenocarcinoma Using High-Density Single-Nucleotide Polymorphism Arrays* (Nancarrow, 2008) provides survival data for the patients in the study. However, no analysis was performed to find potential relationships between copy number events and the survival time. Here we report on the findings of such correlations discovered using the Predicative Power feature in Nexus Copy Number.

The Cancer Survival Data

The cancer survival data are provided by Nancarrow et. al. in their paper *Genome-Wide Copy Number Analysis in Esophageal Adenocarcinoma Using High-Density Single-Nucleotide Polymorphism Arrays* (Nancarrow, 2008), in which Illumina Sentrix HumanHap300 BeadChips were used. The BeadStudio processed data set containing log ratios and B-allele frequencies was downloaded from GEO (Gene Expression Omnibus, www.ncbi.nlm.nih.gov/geo) and directly loaded into Nexus Copy Number. A total of 23 samples with survival time and other sample factors are displayed in the Data Set tab (Figure 1). The copy number segmentation of all samples was done using BioDiscovery's proprietary SNPRank Segmentation algorithm in Nexus Copy Number.

The copy number events of each sample contribute to the frequency plot shown in Figure 2. Peaks in green (copy number gains) and in red (copy number losses) indicate percentages of total samples with the same copy number gains or losses in that region. The lower part of the window displays the copy number events in each sample.

Nexus - Esophageal Cancer Project (Human NCBI Build 36.1)

File Help

Data Set Comparisons External Data Results

Load Descriptor Load Data View Select Delete Reset Add Factor

Sample	Data Type	Status	Quality	Gender	Stage	Diagnosis	report	Survival	Event	tumor	Control	Gender	GSM	Tumor % by LOH only	Age at diagnosis
40359	Illumina	Processed	0.151	M	I	Slide	review	469	Alive	EAC	F		GSM266119-6630	58	55
40363	Illumina	Processed	0.128	M	I	Pathology	report	462	Alive	EAC	F		GSM266661-7759	78	69
40320	Illumina	Processed	0.17	M	II	Slide	review	699	Alive	EAC	F		GSM265808-5854	87	62
40325	Illumina	Processed	0.232	M	II	Slide	review	657	Dead	EAC	F		GSM266074-6524	84	62
40334	Illumina	Processed	0.154	M	II	Slide	review	607	Alive	EAC	F		GSM265786-5796	78	68
40360	Illumina	Processed	0.153	M	II	Slide	review	452	Alive	EAC	F		GSM266707-7836	89	78
42199	Illumina	Processed	0.186	F	III	Pathology	report	634	Alive	EAC	F		GSM266586-7268	65	74
53048	Illumina	Processed	0.132	M	III	Pathology	report	1153	Alive	EAC	F		GSM266703-7832	67	70
54043	Illumina	Processed	0.181	M	III	Pathology	report	1130	Alive	EAC	F		GSM265501-5275	63	73
40331	Illumina	Processed	0.151	M	III	Pathology	report	600	Alive	EAC	F		GSM266075-6525	63	65
40341	Illumina	Processed	0.171	M	III	Slide	review	391	Dead	EAC	F		GSM265789-5799	78	79
40345	Illumina	Processed	0.182	M	III	Slide	review	547	Dead	EAC	F		GSM265790-5800	91	61
40353	Illumina	Processed	0.128	M	III	Slide	review	536	Alive	EAC	F		GSM265791-5801	54	63
40356	Illumina	Processed	0.131	M	III	Slide	review	489	Dead	EAC	F		GSM266705-7834	69	78
40361	Illumina	Processed	0.155	M	III	Slide	review	282	Dead	EAC	F		GSM266659-7757	67	75
40323	Illumina	Processed	0.201	M	IV	Slide	review	531	Dead	EAC	F		GSM265809-5855	51	60
40338	Illumina	Processed	0.134	M	IV	Slide	review	53	Dead	EAC	F		GSM265787-5797	85	72
40340	Illumina	Processed	0.132	M	IV	Slide	review	329	Dead	EAC	F		GSM265788-5798	59	66
40358	Illumina	Processed	0.123	M	IV	Slide	review	408	Dead	EAC	F		GSM266706-7835	58	56
40362	Illumina	Processed	0.146	M	IV	Slide	review	183	Dead	EAC	F		GSM266660-7758	61	65
40364	Illumina	Processed	0.166	M	IV	Slide	review	287	Dead	EAC	F		GSM266708-7837	90	71
53145	Illumina	Processed	0.128	M	IV	Pathology	report	878	Alive	EAC	F		GSM266715-7845	88	52
54014	Illumina	Processed	0.168	M	IV	Pathology	report	367	Dead	EAC	F		GSM265500-5247	67	80

Selected Samples: 23

Fig. 1. A total of 23 samples were imported and processed in Nexus Copy Number. QC scores were calculated for the data qualities. Multiple sample factors, including Survival, were added to describe the sample attributes.

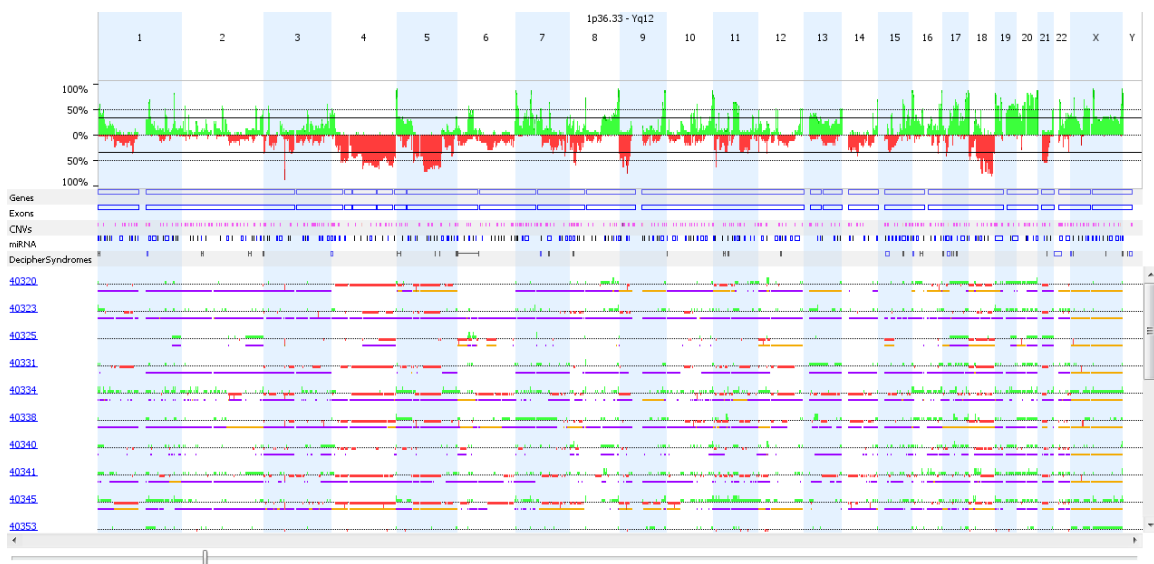


Fig. 2. Copy Number events for all samples are represented in the frequency plot at the top, with copy number gains in green and copy number losses in red. The CNV profiles for individual samples are shown at the bottom.

Kaplan-Meier Plot

The Kaplan-Meier plot is a common method used in medical research to describe survival characteristics of a group of patients and is typically used to compare survival times between different groups. For example, the Kaplan-Meier plot of the patients in this study shows a strong correlation between cancer stages and survival of the patients (Figure 3). Since different copy number events exist in different cancer stages, it is reasonable to think that a correlation between copy number events and patient survival time must also exist.

The goal of this paper is to determine if such correlation can be established, as illustrated in Figure 4, and how to locate these regions with varying copy number events.

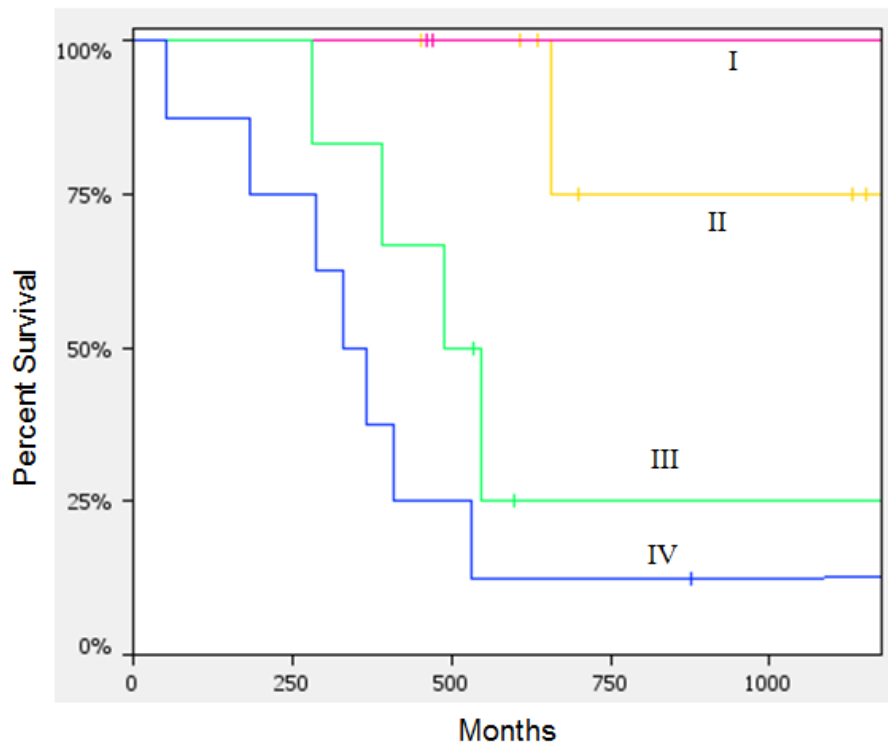


Fig. 3. Kaplan-Meier plot in Nexus Copy Number shows a strong correlation between survival and cancer stages.

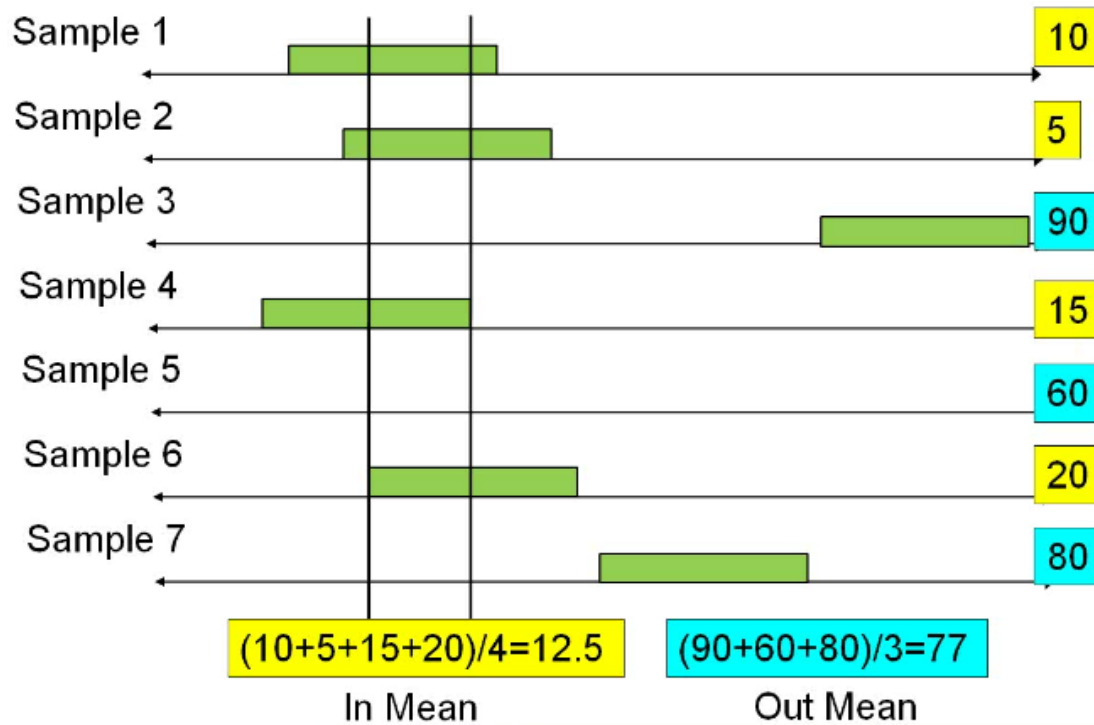


Fig. 4. Schematic illustration of the relationship between copy number events and survival time. In this example, patients with copy number gain in a certain region have shorter survival while patients without this event lives longer as indicated by the mean survival time with or without this event (In Mean and Out Mean, respectively).

Predicative Power

The Predicative Power feature generates the results for such data analysis in an easy and straightforward fashion. Looking at only the **Alive** patient samples with survival time longer than 600 days and at all the **Dead** patient samples, the Predicative Power function locates the CNV regions that correlate to the survival time (with a p-value cut-off of 0.01) and displays them in a table (Figure 5). Take region chr1:224,224,526-224,921,488 as an example, it has a total of 8 genes displaying a copy number gain. Of all the samples being analyzed, 15 samples with this event have a mean survival time of 449.2 days while 2 samples without this event have a mean survival time of 1,141.5 days (Figure 6).

Region	Event	Length	# Genes	In	In Mean	Out	Out Mean	Stdev	P-Value
chr1:224,224,526-224,921,488	CN Gain	696,963	8	17	467.176	2	1,141.5	190.877	0.006
chr1:224,921,488-225,012,149	CN Gain	90,662	0	16	465.812	3	924	226.597	0.009
chr7:148,634,639-150,067,660	CN Gain	1,233,022	19	5	248.6	14	641.571	222.072	0.006
chr7:150,825,173-151,006,173	CN Gain	181,001	0	8	337.375	11	684.182	223.475	0.004
chr7:151,015,079-151,018,964	CN Gain	3,886	0	6	288.667	13	653.308	224.791	0.006
chr11:60,285,142-60,310,850	CN Gain	25,709	0	9	369.444	10	690	231.608	0.01
chr11:60,310,850-60,321,974	CN Gain	11,125	0	10	381.4	9	712.333	227.94	0.007
chr11:63,458,720-65,362,930	CN Gain	1,904,211	75	11	401.909	8	725.5	231.808	0.01
chr12:123,269,250-123,773,357	CN Gain	504,108	2	12	400.5	7	774.143	216.275	0.002
chr15:70,808,672-71,347,272	CN Gain	538,601	1	2	1,141.5	17	467.176	190.877	0.006
chr15:73,896,428-75,576,660	CN Gain	1,680,233	14	2	1,141.5	17	467.176	190.877	0.006
chr15:77,295,470-82,987,545	CN Gain	5,692,076	50	3	943.333	16	462.188	220.179	0.007
chr15:91,601,453-92,174,029	CN Gain	572,577	0	5	815	14	439.286	227.789	0.009
chr15:92,174,029-92,495,000	CN Gain	320,972	0	4	882	14	446.467	218.478	0.005
chr16:83,656,267-83,676,307	CN Gain	20,041	0	10	378.5	9	715.556	225.692	0.005
chr16:84,244,887-84,264,346	CN Gain	19,460	0	12	392.25	7	788.286	206.796	0.001
chr16:84,264,346-84,943,218	CN Gain	78,873	2	10	384.7	9	708.667	230.422	0.009
chr17:45,274,568-46,246,419	CN Gain	971,852	25	12	410.417	7	757.143	226.465	0.007
chr17:71,080,297-71,669,276	CN Gain	588,980	23	15	446.733	4	881	218.897	0.005
chr17:71,866,247-74,317,788	CN Gain	2,451,540	34	15	446.733	4	881	218.897	0.005
chr17:74,317,786-74,438,247	CN Gain	120,462	1	15	446.733	4	881	218.897	0.005
chr17:74,438,247-74,529,819	CN Gain	91,573	2	16	456.75	3	972.333	209.558	0.004
chr17:76,135,463-76,426,025	CN Gain	290,564	0	16	456.75	3	972.333	209.558	0.004
chr18:33,035,939-33,056,466	CN Gain	20,528	0	4	888	15	444.867	215.923	0.004
chr20:43,150,724-43,348,330	CN Gain	197,607	7	16	456.75	3	972.333	209.558	0.004
chr20:43,348,330-43,507,436	CN Gain	159,107	8	16	456.75	3	972.333	209.558	0.004
chr20:43,507,436-49,755,719	CN Gain	6,248,284	73	16	456.75	3	972.333	209.558	0.004
chr20:49,755,719-49,797,199	CN Gain	41,481	0	15	447.2	4	879.25	219.625	0.005
chr20:54,575,636-59,565,684	CN Gain	4,989,949	32	16	456.75	3	972.333	209.558	0.004
chr20:59,565,684-59,585,010	CN Gain	19,327	0	16	456.75	3	972.333	209.558	0.004
chr4:130,143,922-133,890,061	CN Loss	3,746,140	2	14	437.5	5	820	225.589	0.007
chr4:133,890,061-134,066,461	CN Loss	176,401	0	13	422.385	6	789	224.096	0.006
chr4:139,395,522-139,570,274	CN Loss	174,753	0	14	439.143	5	815.4	227.616	0.009

Fig. 5. Predictive Power function quickly locates the regions that show significant correlations between copy number events and survival.

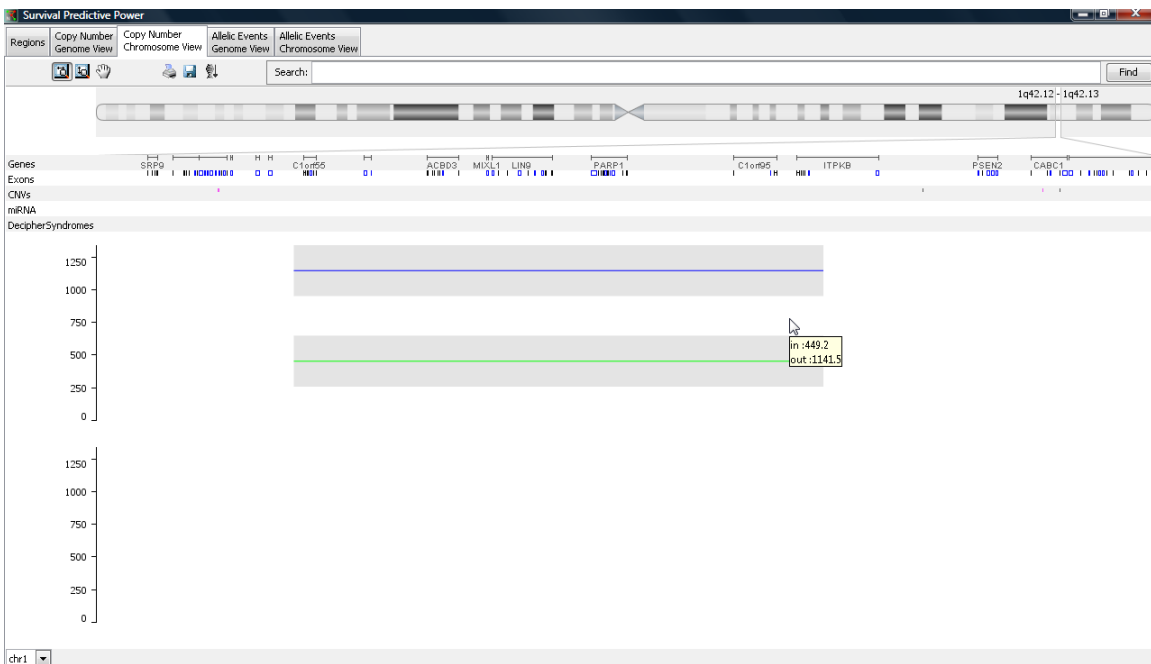


Fig. 6. A copy number gain region on Chromosome 1 with significant correlation with shorter survival. Copy number gain is shown in green and no event in blue. The two gray areas indicate the ranges of standard deviations of survival in the two patient groups.

Conclusion

The Predicative Power feature in Nexus Copy Number provides an easy way for researchers to locate the CNV regions that are highly correlated to cancer patient survival time. Along with a Kaplan-Meier plot, as well as other useful features, Nexus Copy Number can be a powerful yet user-friendly tool for scientists in cancer CGH research. Analysis of the data by scientists themselves who are the experts in their fields rather than by external bioinformaticians gives the scientists complete control over posing relevant biological questions and quickly uncovering answers to such questions.

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