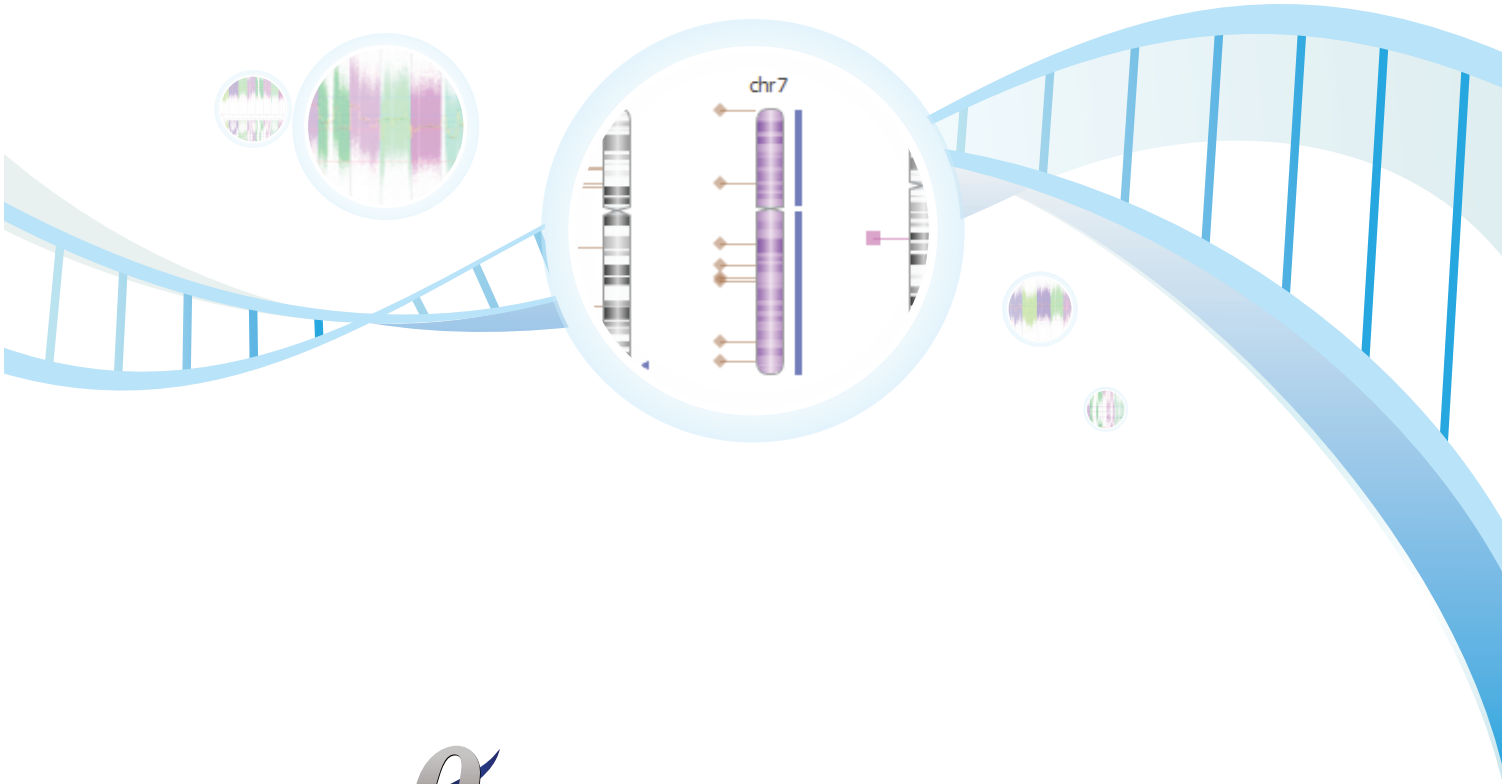
TCGA PREMIER

Curated CNVs, Sequence Variation, Simplified Access



BioDiscovery

SOFTWARE SOLUTIONS FOR LIFE SCIENCES

5155 Rosecrans Avenue, Suite 310 - Hawthorne, CA 90250
Tel: +1 310-414-8100 / URL: www.biodiscovery.com / Email: info@biodiscovery.com

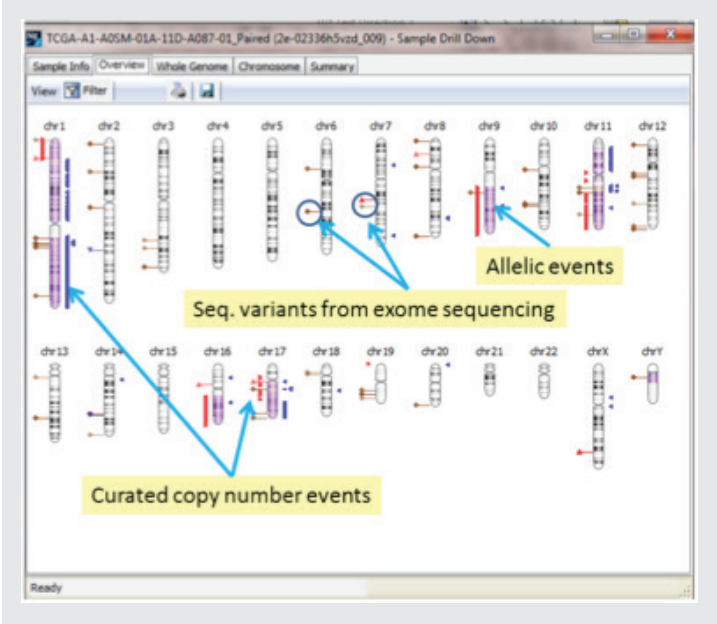
www.biodiscovery.com

TCGA PREMIER

Curated CNVs, Sequence Variation, Simplified Access

The Cancer Genome Atlas (TCGA) is an amazing resource, containing genomic profiles of thousands of tumors across more than thirty cancer types. The TCGA project has already proven useful in large-scale studies; however, these early studies have principally found the "low hanging fruits", and the voluminous data generated by this project has great additional potential.

The initial processing of the samples produced results suitable for population-wide profiles, but left individual samples over-segmented and many having incorrect ploidy. By correcting these, and integrating sequence variants, TCGA Premier enables the data set to become more useful for studies such as marker discovery, survival analysis, and tumor sub-type profiling.



BRCA sample with curated copy number and sequence variants shown

The Approach

Beginning with the raw data available through special access, we have paired the tumors with their matched normal to remove germline polymorphisms. Expert scientists then review the genomic profiles of each sample, correct for baseline ploidy and segmentation, and manually call copy-neutral homozygous segments (LOH) and events on the mosaic threshold boundary. Sequence variant information is integrated from exome sequencing data, and the resultant dataset made available through Nexus Copy Number.

Figure 1 - a BRCA sample prior to review

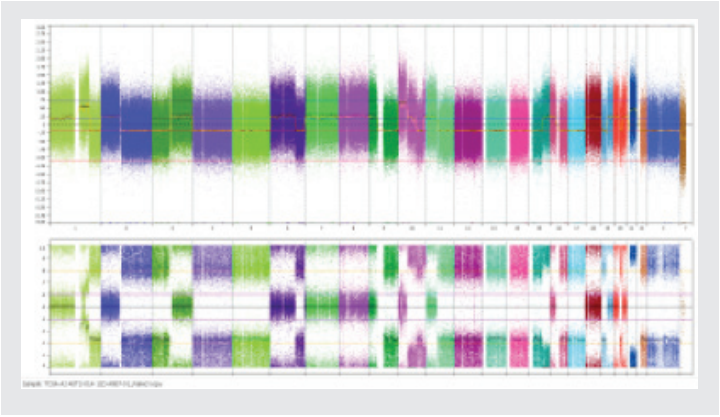


Figure 2 - same BRCA sample after curating step

To further enhance usability, samples with complex tumor profiles are "flagged" for the researcher, as are suspected chromothriptic events. Lastly, samples with poor quality that introduce statistical noise are removed.

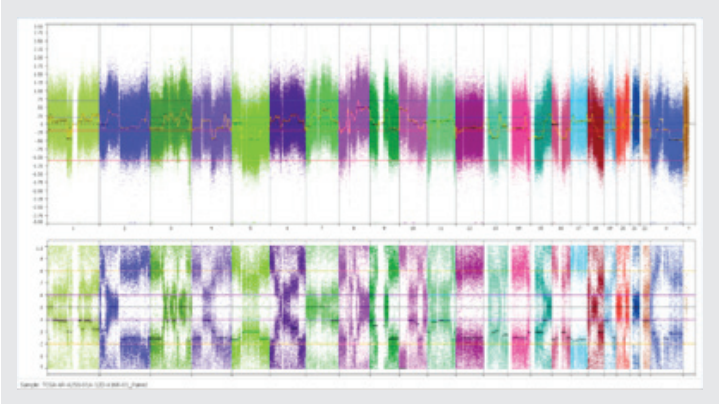


Figure 3 - Example of a "complex" genomic profile

The data set is fully annotated with reviewer comments for specific samples, calling parameter settings that were used, other sample-specific information along with detailed clinical annotations, such as survival, tumor grade, etc., that foster further analyses.

Nexus Interface to TCGA

Access to the data is through Nexus Copy Number software clients. Users can see all data sets, and query or download with mouse-clicks, as well as perform downstream analyses. There is no programming or scripting involved.

Example Results

Illustrating improvements in the copy number profiles – these two CNV frequency plots were selected from the breast invasive ductal carcinoma (BRCA) data set. In both cases the same samples are selected which are annotated as ErbB2+ via a FISH assay.

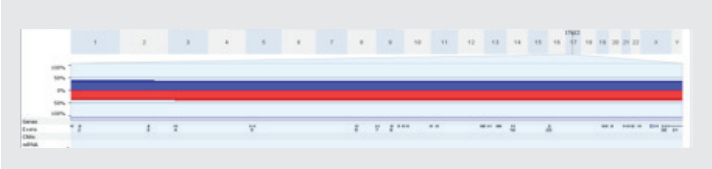


Figure 4 - from TCGA "level 3"
<50% positive and many ErbB2 "losses" for FISH positive samples

After review of more than 800 samples, re-analysis shows results more consistent with the expected FISH results. The difference between 100% concordance is possibly due to CEP17 ratio skewing in the FISH assay[1].

Figure 5 - re-analysis of curated data shows 85% with positive ErbB2 status, and 0% with copy number loss

Rb1 shows a similar difference between the curated and TCGA "level 3" data. While the un-curated "level 3" data show a rb1 loss of 44% and copy number gains in 10% of the BRCA population , the curated TCGA Premier reveals copy number losses of rb1 of 27%, consistent with the literature[2].

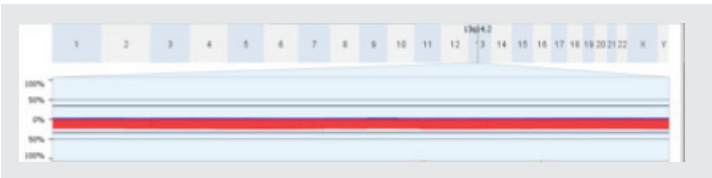
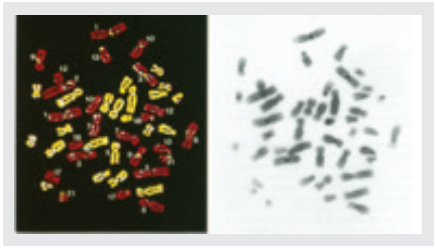


Figure 6 - rb1 loss of 27% in re-analyzed data, consistent with literature consensus[2]

In the kidney chromophobe data set (KICH), the effect of correct ploidy assignment and re-segmentation are readily visible. In the Level 3 data, copy number gains are assigned to a large fraction of the samples, across several chromosomes, which are inconsistent with the expected findings for this cancer type[3]. By contrast, the re-analysis revealed profiles in line with expectations which will be more useful for further analysis, use in other studies, and applications for validation.

Figure 7

KICH Karyotype showing hypodiploid[3]



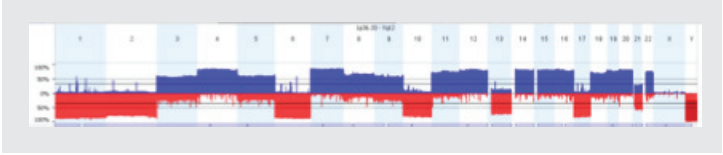


Figure 8 - KICH level 3 data showing CNV gains across several chromosomes inconsistent with the literature[3]

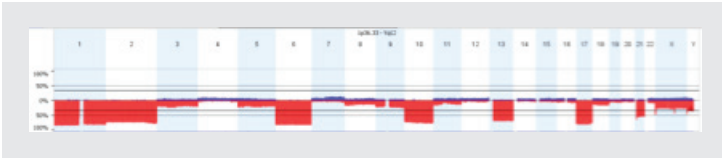


Figure 9 - Re-analyzed KICH data showing characteristic hypodiploid profile

References

1) Gunn S, Yeh I-T, Lytvak I, et al. Clinical array–based karyotyping of breast cancer with equivocal HER2 status resolves gene copy number and reveals chromosome 17 complexity. BMC Cancer. 2010;10:396.doi.

2) Robinson, T. J., Liu, J. C., Vizeacoumar, F., Sun, T., Maclean, N., Egan, S. E., ... & Zacksenhaus, E. (2013). RB1 Status in Triple Negative Breast Cancer Cells Dictates Response to Radiation Treatment and Selective Therapeutic Drugs. PloS one, 8(11), e78641.

3) Speicher, M. R., Schoell, B., du Manoir, S., Schröck, E., Ried, T., Cremer, T., ... & Kovacs, G. (1994). Specific loss of chromosomes 1, 2, 6, 10, 13, 17, and 21 in chromophobe renal cell carcinomas revealed by comparative genomic hybridization. The American journal of pathology, 145(2), 356.

TCGA PREMIER

Curated CNVs, Sequence Variation, Simplified Access

Improved Copy Number Profiles

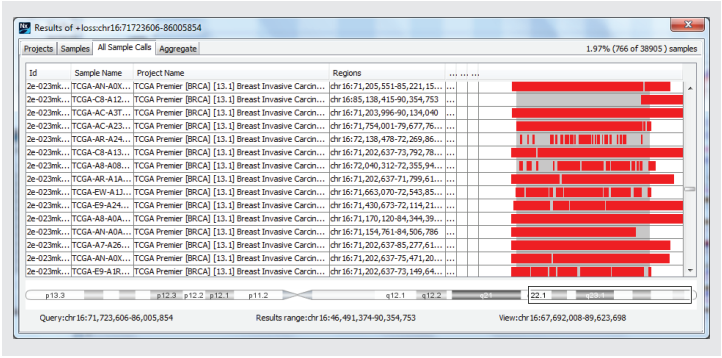
- Each sample is re-processed from the "raw" (i.e., level 1) data, then manually reviewed by a team of scientists ensuring the highest fidelity CNV calls.

Integrated Sequence Variants

- Sequence variants obtained by whole-exome sequencing, providing a more comprehensive genomic picture.

Proven Nexus Copy Number Interface to the TCGA, or use your own pipeline.

- Query by gene, region, an event, or clinical annotation – typically in seconds. Data can be downloaded and integrated with other sources for more robust analyses.

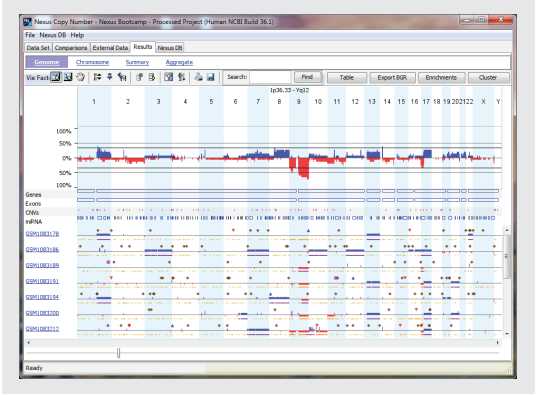


Query of all TCGA BRCA samples for a common deletion.

NEXUS COPY NUMBER

Long the standard for CNV analysis, in use at the world's largest cancer research centers with thousands of users and cited in hundreds of papers, Nexus Copy Number is comprehensive and has been optimized for ease-of-access for TCGA Premier.

- Computation of sample ploidy and tumor burden
- Detection of low-level mosaic events
- Survival analysis and predictive statistics
- Easy statistical and visualization tools



BioDiscovery
SOFTWARE SOLUTIONS FOR LIFE SCIENCES

5155 Rosecrans Avenue, Suite 310 - Hawthorne, CA 90250
Tel: +1 310-414-8100 / URL: www.biodiscovery.com / Email: info@biodiscovery.com
www.biodiscovery.com