

# CROSS PLATFORM ARRAY CGH ANALYSIS

TOPICS IN COPY NUMBER VARIATION ANALYSIS

## CONTENTS

Introduction.....	2
Solution .....	2
Copy Number Variation: Two Datasets .....	2
References.....	5
Contact.....	6

## INTRODUCTION

Array CGH and other copy number detection techniques are becoming very popular among specialists in human genetics and cytogenetics. Related data analysis methods increase in their complexity as datasets rapidly increase in their size.

Bioinformatics methods specifically designed to handle complex experiments involving hundreds of arrays are becoming widely employed. Those methods should among other issues address compatibility of data sets extracted from different array platforms.

## SOLUTION

“CGH Fusion™” is a leading commercially validated package for analysis of array CGH data across multiple samples. It is capable of building aberration frequency profiles for multiple samples within a specific biological context. CGH Fusion facilitates creation and comparison of copy number variation profiles independently of the actual array technology and design used in biological experiments.

The main advantage of CGH Fusion is its ability to fuse data from different array CGH platforms in a single comprehensive view for an instant cross-platform analysis or comparison. To illustrate this ability we have analyzed several publicly available datasets below

To download a 14 day trial of CGH Fusion visit [www.infoquant.com](http://www.infoquant.com)

See References for a complete list of datasets and links to public downloads

## COPY NUMBER VARIATION: TWO DATASETS

Multiple studies targeting discovery of copy number variation in healthy individuals were recently performed by various institutions worldwide. Comparison and utilization of their results at the aberration track level are facilitated by publicly available

knowledge databases like “Database of Genomic Variants”, “Ensemble Genome Browser”, “UCSC Genome Browser” etc. However, we would like to demonstrate how CGH Fusion facilitates comparative analysis of such datasets at the level of actual aberration frequencies. We picked just two of publicly available datasets for our demonstration.

- The Copy Number Variation (CNV) project involves multiple institutions and is led by Sanger Institute (UK). 270 seemingly healthy individuals using samples from HapMap project were studied for chromosomal regions of high genomic variation using Sanger WGTP arrays. Raw data files in BlueFuse format can be accessed from the project website.
- Similar research was led by the Imperial College London (UK) and Agilent Technologies (U.S.). This particular study involved samples from 50 healthy males and was performed using custom Agilent 244K arrays. Log-ratio data is provided in Gene Expression Omnibus series matrix format (accession number GSE8691).

Measurement data from both studies were loaded into CGH Fusion for simultaneous analysis. For every array the following analysis steps were performed within the automated processing pipeline.

- Data normalization using array CGH adopted version of LOWESS algorithm (performed for Sanger data only).
- Combining replicate measurements from two dye-swap arrays for every clone (performed for Sanger data only).
- Detection of chromosomal aberration (gain/loss) regions using our robust algorithm for ratio data segmentation.

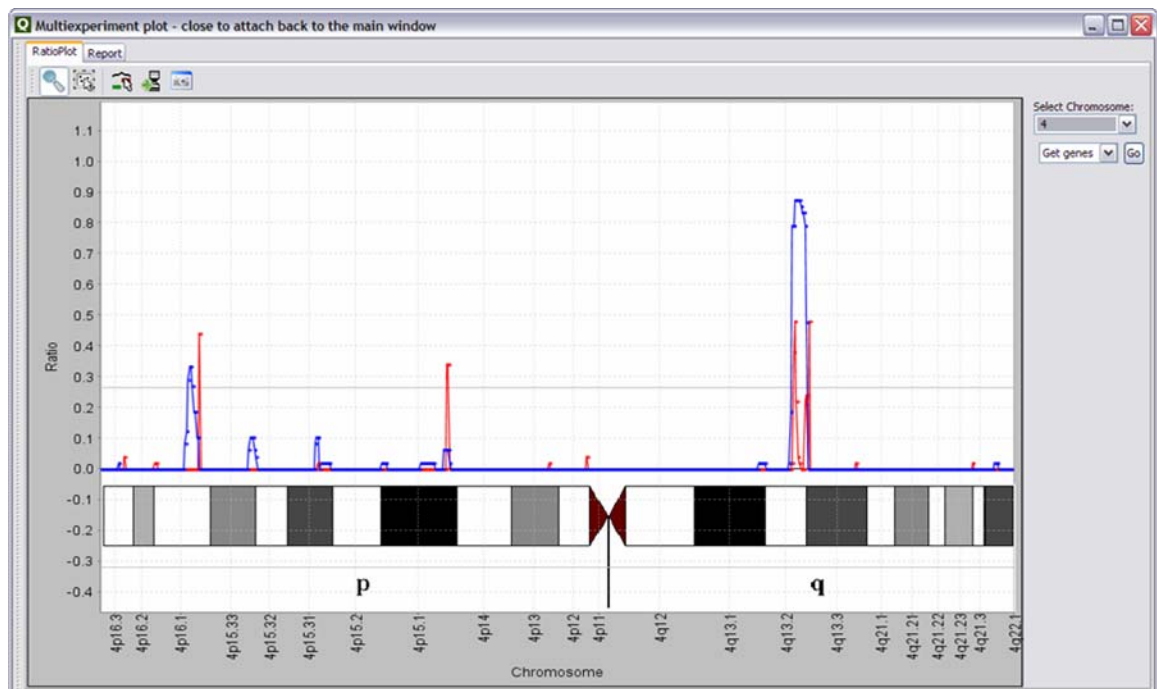
Data fusion between samples within each study and across both studies was performed in two steps.

- In each of the two datasets the software assessed every clone’s frequency of being marked as an aberration across arrays. For instance, if a clone belongs to an aberration region in 10 out of 50 patients, its aberration frequency would be 0.2 (20%).

- Aberration frequency profiles were re-computed based on a common spatial grid in terms of chromosomal position for both platforms. This step insures that aberration frequencies are shown in a manner consistent between different array platforms.

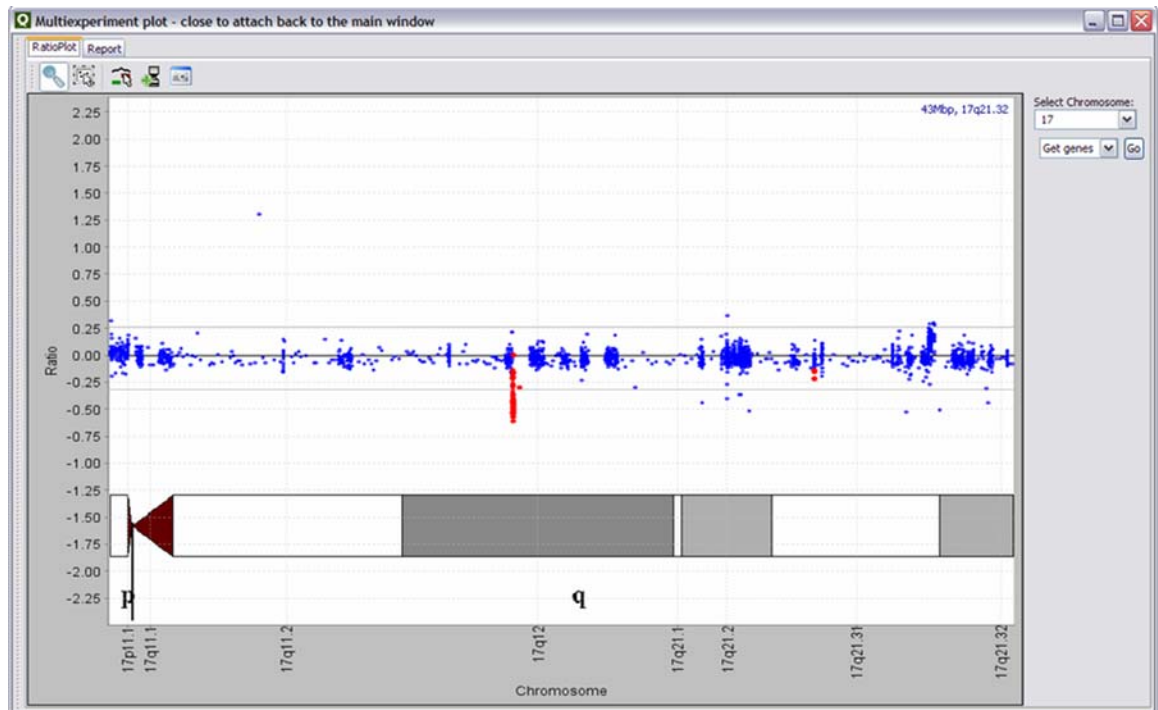
Genomic variation profiles in terms of aberration frequencies arranged along the genome are presented for both studies on visualization plots below. Sanger data are shown in blue color and Agilent 244K array aberration frequencies are drawn in red.

Significant portions of aberration regions show concordance across both platforms. You can see a good example of that on the screenshot for chromosome 4. Blue and red peaks overlay quite well.



However, there are some differences in detected regions between the studies as well. Agilent arrays were able to detect more micro-regions of active genomic variation than Sanger arrays. That can be explained by higher resolution of Agilent 244K arrays. At the same time Sanger arrays detected some regions of high genomic

variance that Imperial College's experiments missed. This might be due to the fact that Agilent arrays were specifically customized to show higher density around regions that the project group suspected to have high variance from their prior experience. Therefore arrays studied by Imperial College demonstrated lower resolution outside those regions. To illustrate uneven spacing of the probes, we show a log-ratio plot for a part of chromosome 17 for one of the arrays from the study by Imperial College below.



We conclude that a combination of CNV studies done using different array CGH platform may provide fuller information about regions of genomic variance.

## REFERENCES

- A. J. de Smith , A. Tsalenko , N. Sampas , A. Scheffer , N. A. Yamada , P. Tsang , A. Ben-Dor , Z. Yakhini , R. J. Ellis , L. Bruhn , S. Laderman , P. Froguel , and A. IF. Blakemore, "Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for

association studies of complex diseases”, Human Molecular Genetics, Advance Access published on July 31, 2007

B. The Copy Number Variation (CNV) Project:

<http://www.sanger.ac.uk/humgen/cnv/>

## CONTACT

For additional information regarding CGH Fusion, the topics discussed in this paper or reproduction rights, please contact infoQuant, Ltd. ([support@infoquant.com](mailto:support@infoquant.com)) or visit [www.infoquant.com](http://www.infoquant.com).

*CGH Fusion is a trademark of infoQuant, Ltd. Other products, names and marks may be trademarks of their respective organization. infoQuant makes no guarantee that the materials contained within this document are error-free and reserves the right to update the contents at anytime.*

