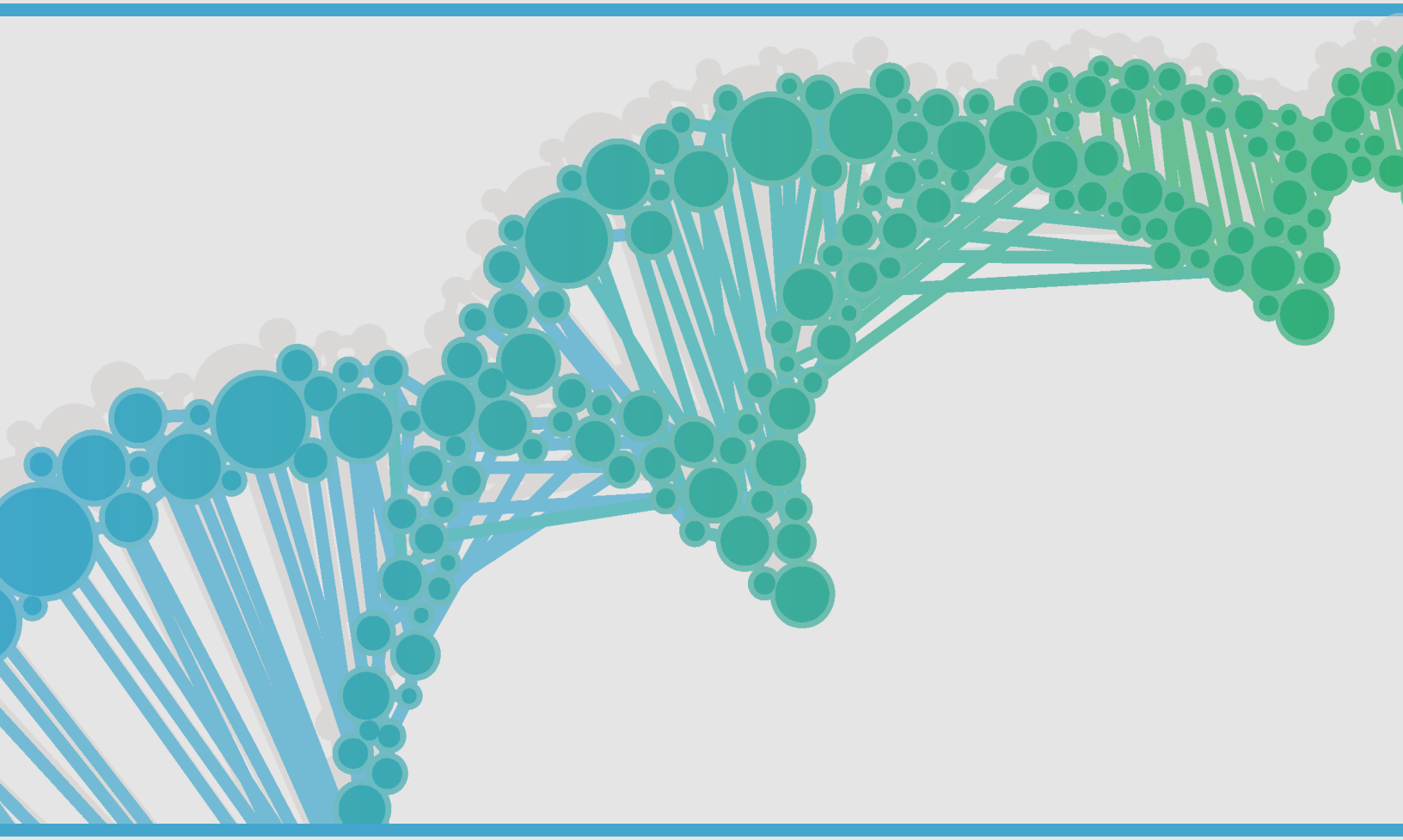
BioDiscovery



WHITEPAPER

Addressing Aneuploidy and Nonaberrant Cell Admixture in Tumor Samples for Copy Number Variation (CNV) and Loss of Heterozygosity (LOH) Analysis

Explore the use of LogR and B-allele frequency values from SNP arrays to interpret the underlying biology of tumor samples, along with limitations of this approach.

www.BioDiscovery.com

White Paper: Addressing Aneuploidy and Nonaberrant Cell Admixture in Tumor Samples for Copy Number Variation (CNV) and Loss of Heterozygosity (LOH) Analysis

Louis Culot, Zhiwei Che, Shalini Verma, August 10, 2012

Summary

Tumor samples, especially solid tumors, often contain a mixture of both aberrant tumor cells and non-aberrant normal cells. This white paper explores the use of LogR and B-allele frequency values from SNP arrays to interpret the underlying biology of tumor samples, along with limitations of this approach. It also reviews the ASCAT algorithm and its use when matched-normal samples are available.

LogR and B-allele Frequency

The Log Ratio (LogR) at a particular position on the genome is calculated by using a probe to take the signal intensity of the sample at that position and dividing it by the signal intensity for a reference sample at the same position. Using “2-color” array technology, the same probe is used to simultaneously measure the signal for the test and reference samples using a competitive hybridization approach. Many arrays, including most SNP array platforms, use a single color approach. In this case, the reference DNA signal is provided ‘off chip’, and the data is normalized to this off-chip reference. Since there can be systematic biases coming from measurements made of the sample vs. the reference (e.g., a brighter laser is used for one channel vs. the other), a normalization step is necessary. One common method for this normalization is to adjust the LogR value of the median probe to zero (so there would be equal number of probes above and below the zero line). This method works quite well if the there are few and small aberrations in the test genome. By centering the diploid region of the sample at “0” on the LogR scale, copy number gains or losses can be detected by shifts in amplitude. However, using this value alone has some limitations when there are significantly large genomic changes (as in many cancers and highly aneuploid samples).

Array platforms with SNP probes can be employed to address some of the above limitations. In the example, below, each panel represents a different underlying biological event.

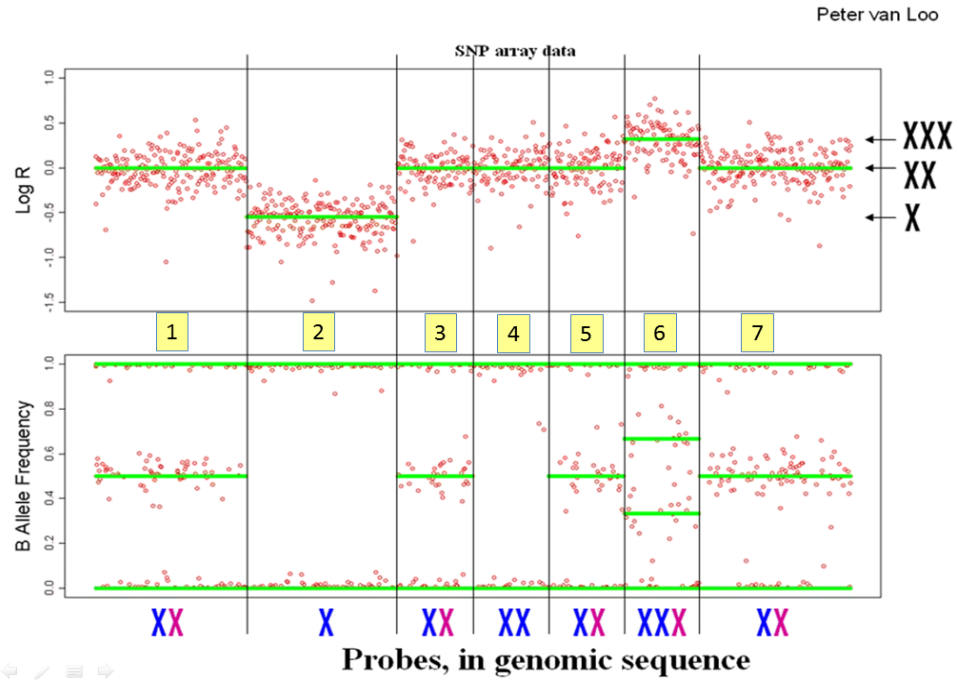


Figure 1

In the left most panel (panel 1) which is a 'normal' example, the LogR is 0 (normalized), and the B-allele frequency (BAF) plot shows a characteristic three-band pattern. The BAF is created from the SNP probes on the array, and represents the frequency of "B" allele (the alleles are typically in alphabetic order, for example, adenosine or cytosine is the "A" allele, while thymine or guanine is the "B" allele). In a diploid heterozygous sample, AA, AB, and BB alleles are present.

The second panel is an example of a single-copy loss. In this case the LogR value is below 0 (the actual amplitude of the shift depends on the dynamic range of the array, scanner, and other inputs prior to the software). The BAF shows only two bands – either "A" only (BAF=0), or "B" only (BAF=1). There are no "AB" calls since the sample has only one allele.

The fourth panel is an example of a disomic homozygous region. This can be created by a number of events, including a copy-neutral loss of heterozygosity where one allele is lost and the remaining allele is replicated. It can also be caused by a *de novo* absence of heterozygosity, such as the result of consanguinity. In this case the LogR value is normalized to 0 (i.e., a 'diploid' result). But the BAF shows a two-band pattern. This is because, from the SNP probe perspective, there is only one allele (there are two identical SNPs with the same chromosome coordinates). So the possible results are either AA (0 on the BAF) or BB (1 on the BAF). There is no "AB" call, which can only occur if a different SNP is present in the same position on both chromosomes.

The sixth panel is showing a single copy gain. This is shown both on the LogR plot (increase in amplitude), and the BAF plot (showing a four-band pattern). The four bands are the possible combination of alleles. Assuming that the alleles are initially heterozygous, the possible alleles will be

AAA, AAB, ABB, and BBB. In figure 2, below, the SNP probes are highlighted in yellow and the resultant calls are shown below.

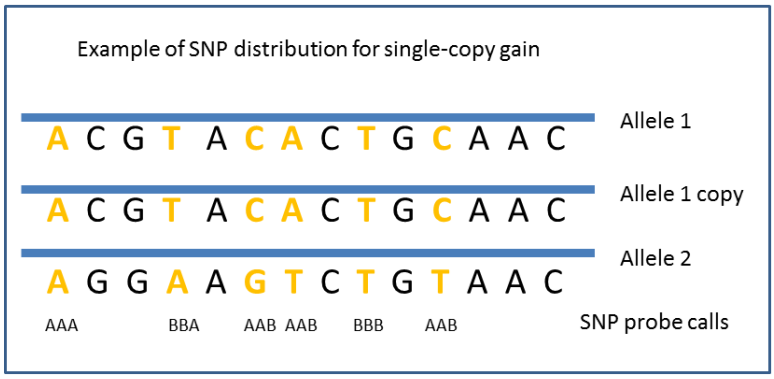


Figure 2

Figure 3, below, is an example of a tetraploid sample. In panel 1, the LogR is 0 and the B-allele frequency shows a three-band pattern. Without any additional information it is not possible to differentiate the example in panel 1 from a diploid sample, since the two-copy gains for the tetraploidy event are balanced (i.e., each allele is replicated once). Likewise panel 2 shows a 2-band B-allele frequency and a LogR of 0, which, in this case, reflects a tetraploid homozygous region. In the absence of other events, panels 1 and 2 would be indistinguishable from a diploid sample with panel 1 being normal and panel 2 being a disomic homozygous region. However, panel 4 gives the clue to what's occurring. In panel 4 the LogR is still 0, but there is a four-band pattern in the BAF. The only explanation for the four-band pattern would be an allelic imbalance (an unbalanced gain). In this case there is a two-copy gain of one allele, and a single copy of the other allele. Looking at panel 4 in isolation also would not reveal the true nature of the sample. One explanation of panel 4 could be a trisomic single-copy gain on a sample whose median probe value is triploid (i.e., the majority of the sample is trisomic). However, if this were the case the three-band pattern in panels 1 and 5 would then see a reduction in amplitude in the LogR (in other words, these would be true 'diploid' regions, but have an artificially low LogR value. Re-centering the logR value in that case would then reflect back in the suspected panel 4 trisomy as a LogR gain). But the LogR in panels 1 and 5 is 0, and accompanied by a three-band BAF pattern. By looking at this sample holistically the only explanation would be a uniform tetraploid event, with a balanced gain in panels 1, 3, and 5, a homozygous stretch across all four alleles in panel 2, and an allelic imbalance (three copies of one allele, one of the other allele) in panel 4.

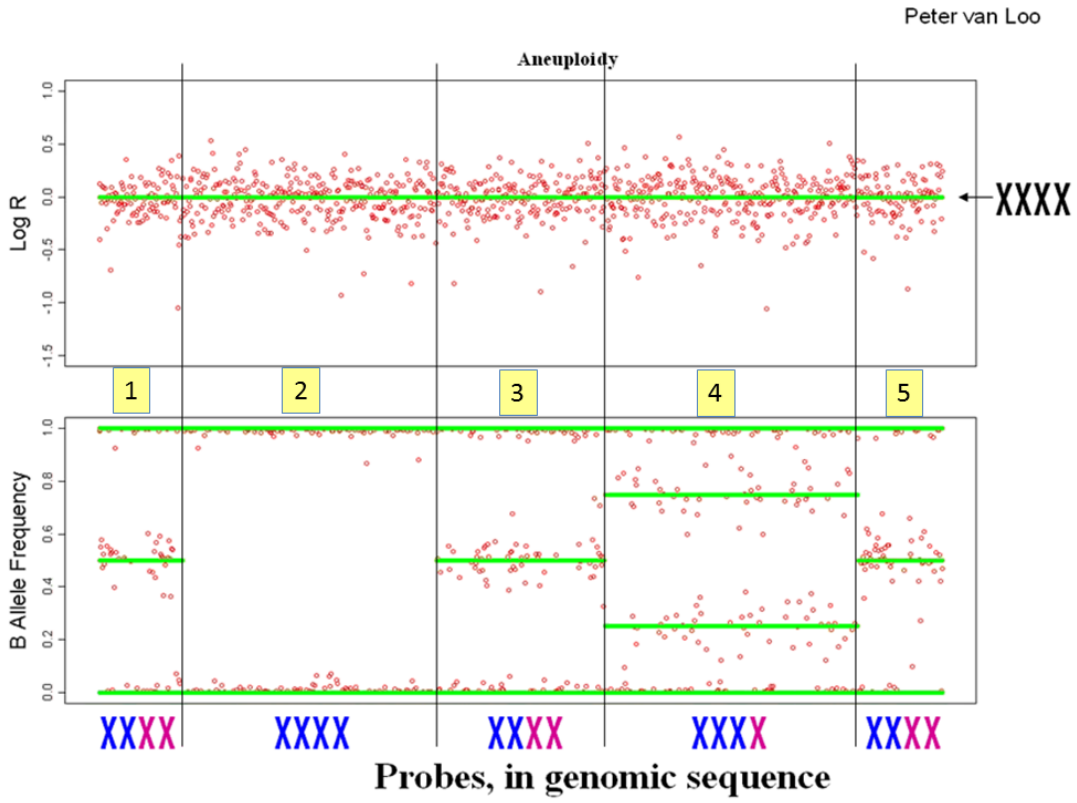


Figure 3

Effect of non-aberrant cell fraction

Samples often contain non-aberrant cells which affect the signal reported by the BAF. In figure 4, below, hypothetical BAF values are shown for a sample that has a mixture of cells (some have a single-copy loss on one chromosome and others are normal).

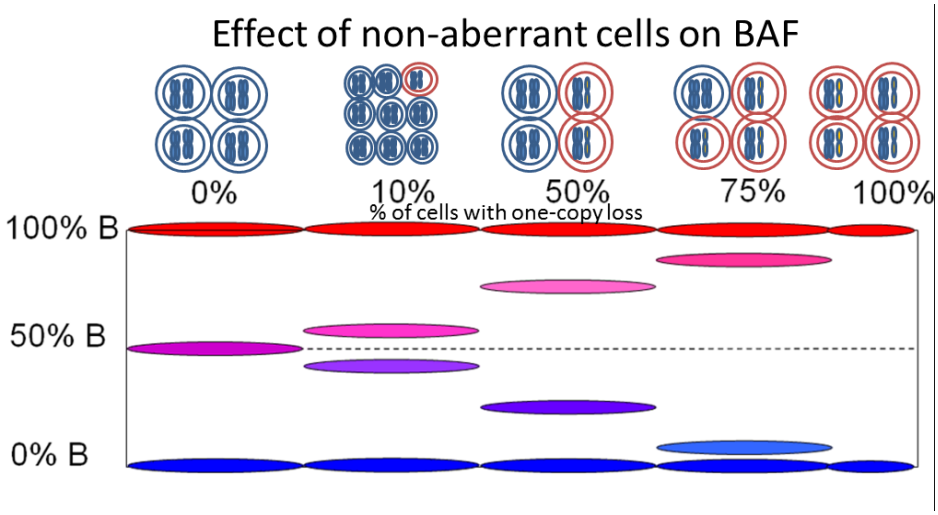


Figure 4

On the far left and right sides of the figure are the BAF results for samples that are either pure normal cells (0% cells with one-copy loss), or pure aberrant cells (100% cells with one-copy loss). Since the signal from the array is from the mixture of DNA from the entire sample, as the number of non-aberrant cells in the sample increases, the percentage of "AB" calls (i.e., 50% "B") also increases, linearly with respect to the non-aberrant cell fraction.

Figure 5 shows a similar example with the LogR value in addition to the BAF:

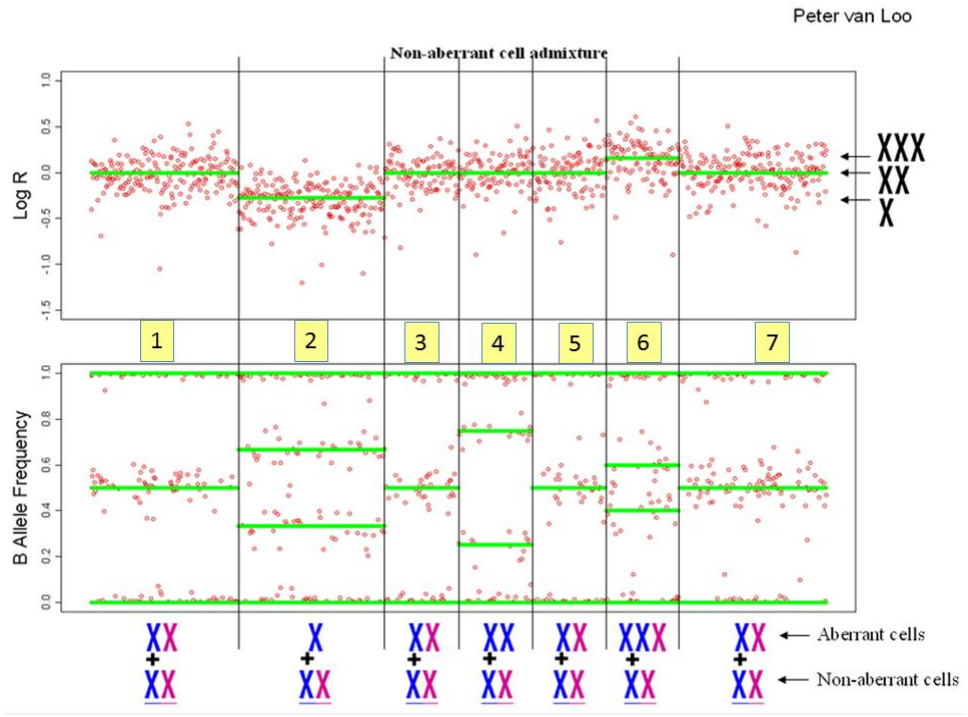


Figure 5

Unlike the pure sample examples, the single-copy loss shown in panel 2 results in a four-band pattern in the BAF. This can be explained by the presence of non-aberrant cells in the fraction. The contribution of the diploid alleles from the chromosome in the normal cell, plus the single allele from the chromosome in the aberrant cell, results in three copies for that same region of DNA. Assuming a 50/50 mixture of tumor and normal cells, then from the BAF for this region alone, it is not possible to distinguish this event from a pure single-copy gain (also three copies from the same region). The reduction in amplitude in the LogR, however, tells us that this is a loss of DNA in that region. Also, the gain shown in panel 6 likewise results in a four-band pattern in the BAF. However, the middle bands are narrower compared to the single-copy loss. This is because the overall presence of the alleles is different. In the case of a gain, there are five total alleles (two from the normal fraction and three from the aberrant fraction). The ratio of the alleles is therefore 3/2. Compared to the single-copy loss where the total number of alleles is three (two from the normal fraction and one from the aberrant fraction), the ratio of the alleles is then 2/1.

The other effect of the non-aberrant cells will be to reduce the amplitude changes in the LogR value for gains and losses. This is again due to the overall sample signal being attenuated by the presence of normal DNA in those regions. The scale of the attenuation will be in proportion to the amount of normal cells in the sample.

ASCAT (Allele Specific Copy Number Analysis of Tumor samples)

This algorithm was developed by Peter Van Loo at the Wellcome Trust Sanger Institute, and is implemented in software packages including BioDiscovery Nexus Copy Number. This method takes both the LogR values and the B-allele frequency values and calculates the copy-number state for each allele in the fraction. From this, the sample can be corrected for ploidy, and the percentage of normal cells in the sample can be deduced. The current implementation of ASCAT requires both a tumor sample and a matched normal sample (from the same individual) to use as a reference.

The following examples from Nexus Copy Number illustrate an actual sample and matched-normal pair using the ASCAT algorithm.

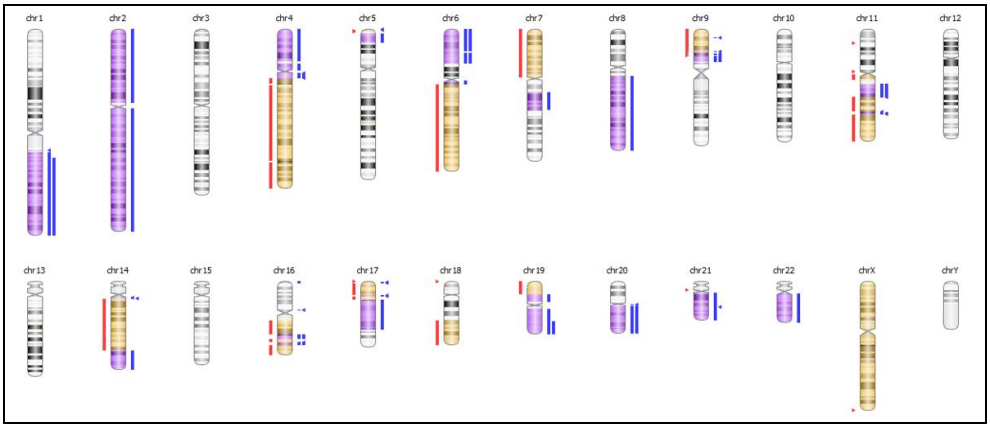


Figure 6

The ideogram in Figure 6 (from Nexus Copy Number) shows the summary results for this sample as processed through the ASCAT algorithm. Drilling down on chromosome 4 shows the effect of the non-aberrant cell fraction on the BAF; note the q-arm deletion, but the 4-band pattern (for ASCAT the software only displays the middle bands – their top and bottom bands are inferred):

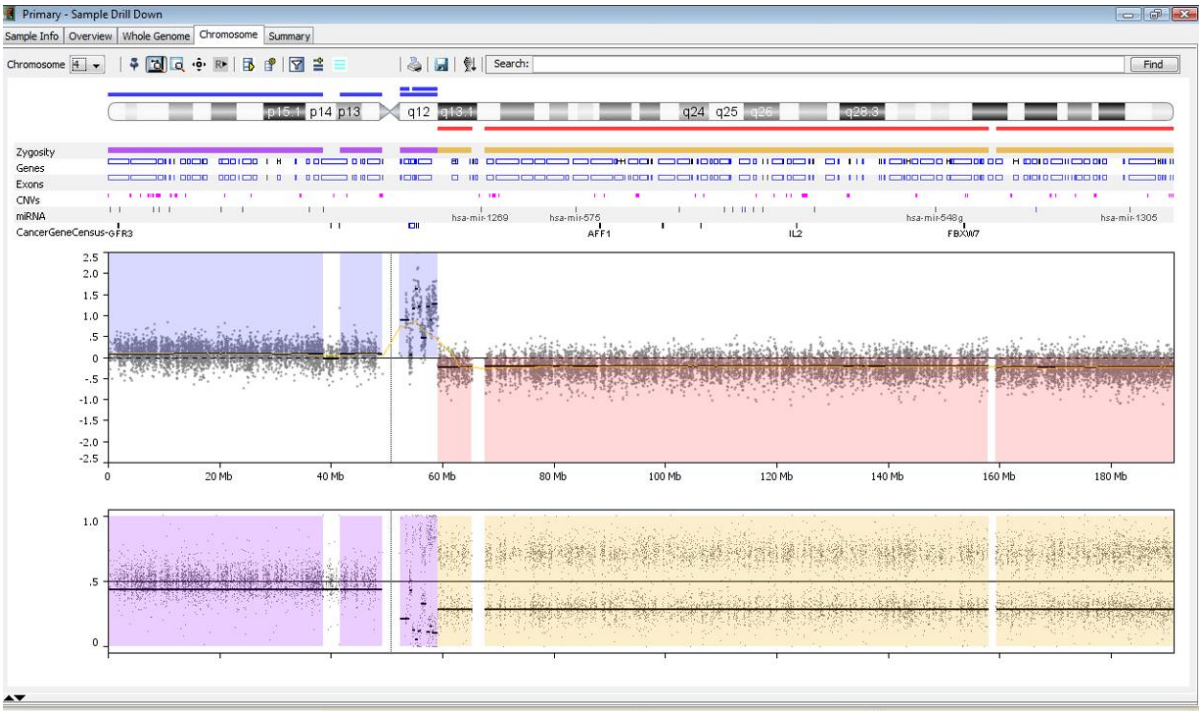


Figure 7

Also of interest is the p-arm amplification, which shows an evolution to hyperploidy near the centromere region, indicated by the increase in amplitude in the LogR plot along with the movement of the BAF middle bands (as the ratio of the amplified allele to the unamplified allele increases, the calls become increasingly "A" or "B").

References

Van Loo et. al. "Allele-Specific Copy Number Analysis of Tumors". Proc Natl Acad Sci USA. 2010 Sep 28;107(39):16910-5

Che, Z. BioDiscovery, Inc. 2012. "Aneuploidy and Nonaberrant cell admixtures in tumor samples for CNV and LOH analysis". [Video File] retrieved from www.biodiscovery.com

BioDiscovery: Nexus Copy Number Discovery Edition. (Version 6.1) [Software]. Available from www.biodiscovery.com

BioDiscovery, Inc. 5155 Rosecrans Avenue, Suite 310 - Hawthorne, CA 90250 USA
Tel: 310-414-8100 - Fax: 310-414-8111 - Email: info@biodiscovery.com - Web: **www.biodiscovery.com**
For Research Use Only