Copy Number Variant Detection Without a Panel of Normals using Anchored Multiplex PCR and Next Generation Sequencing

Taylor R. Patterson, Allison Hadjis, Ryan Rogge, Mark Rogers, Laura Johnson, Devin Tauber | Archer, Boulder, CO. USA.

118

Introduction

Copy number variations (CNVs) are genomic aberrations where the number of copies of a region of the genome differs from the expected number. Errors in DNA replication, repair, or recombination, and other processes can cause CNVs. CNVs may be a cause of disease, a symptom, or both. Copy gains or losses affecting oncogenes or tumor suppressor genes are one mechanism by which cancers may arise, proliferate, or persist. CNVs may be targetable by, or grant resistance to, certain therapies. CNV signatures may indicate chromosomal instability resulting from homologous recombination deficiency (HRD).

CNV methods require a baseline to compare against and calculate fold change values, usually matched normal tissue or a panel of normal samples. Some next-generation sequencing methods allow internal, self-normalization methods but recommend this only for whole genome sequencing data.

Here we introduce a new CNV method which relies only on data from the sample of interest to determine copy number and breakpoints. Our new CNV method is designed to work for IDT Archer VARIANT*Plex*™ and LIQUID*Plex*™ panels and has been released alongside our existing CNV detection method in Archer Analysis version 7.3.

CNV Method Description

Raw reads from a sequenced Anchored Multiplex PCR (AMP™) library are deduplicated to unique fragments using molecular barcodes incorporated during library preparation and aligned to the genome by Archer Analysis as for any DNA workflow.

Next, instead of using a paired normal or panel of normals to normalize counts, a modelling approach is used to account for the influence of common biases (GC%, PCR, sequencing biases, etc.) on observed counts for each primer in the sample. Suspected outliers are removed, then the remaining bias-corrected values proceed to segmentation. The segmentation process groups together adjacent data points of similar value (i.e., copy number).

The sample's baseline copy number is estimated from bias-corrected counts of autosomal primers within the sample. Fold change values are calculated relative to this estimated baseline copy number, rather than a paired normal or a panel of normals. Each segment's mean is tested against the estimated baseline copy number and p-values are reported.

Motivation

CNV methods that require normals, including the existing Archer Analysis CNV method, can produce different results for the same sample library depending on the choice of samples used for normalization (Fig. 1). Input type, quality, read depth, and CNVs present in samples used for comparison can all impact the CNVs that can be detected in an unknown sample. Sourcing, preparing, and analyzing additional samples also increases the amount of work, time, cost to receive results.

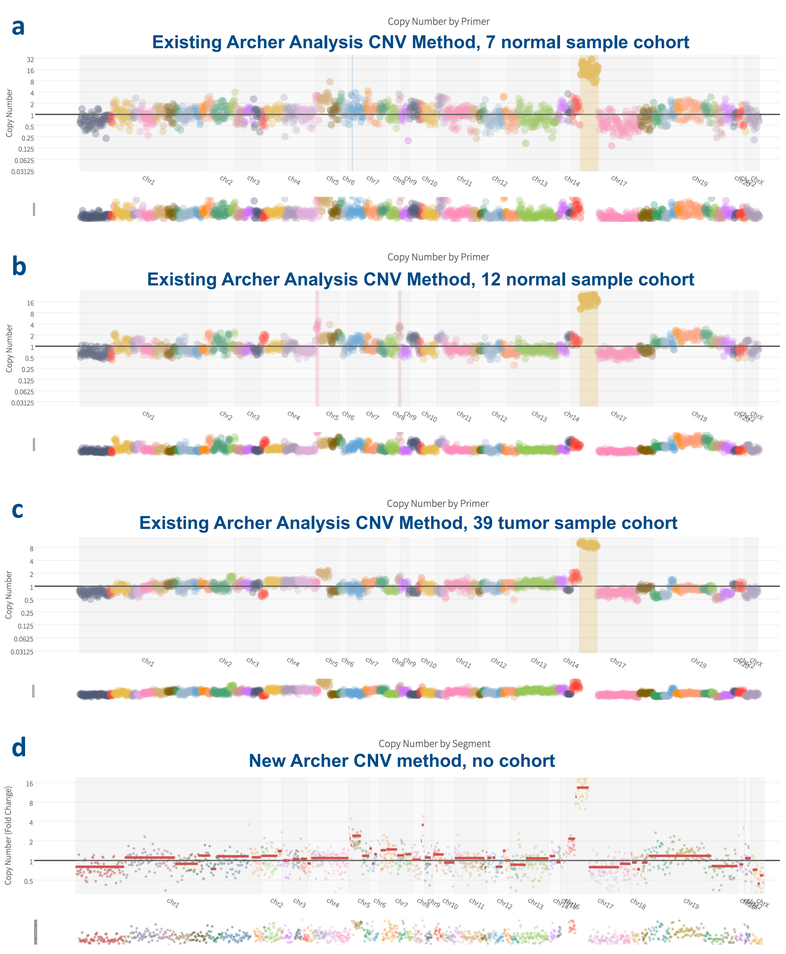


Figure 1. Results of existing Archer Analysis CNV method with three different cohorts and the new CNV method for the same library

50 ng of DNA from FFPE tissue was prepared with our catalog VARIANT*Plex* Pan Solid Tumor panel. Each image in the figure is a cropped screenshot of the Archer Analysis CNV results for this library using our existing method with one of three different cohorts of samples (a, b, c) or the new CNV method that does not require any external samples (d). In each plot, points represent primers. In (d) red bars are CNV segments

In (a), the panel of normals was made up of 7 unrelated normals; a mix of normal adjacent tissues, blood, and reference materials. In (b), the normals consisted of 12 unrelated normal adjacent tissue samples, with no overlap with the normals used in (a). Image (c) is the result of using 39 unknown tumor samples as an approximated 'normal' cohort. The results of the new CNV method which uses only intra-sample data for CNV detection are shown in (d).

In (a,b,c) shaded vertical bars indicate events which pass the default filtering thresholds for the existing method. The new CNV method's visualization (d) does not do this shading.

A large ERBB2 gain and a GNAS gain (not visible in images) are detected in each case, regardles s of the CNV method or normal cohort used. Other than the ERBB2 gain, different genes are highlighted in each of (a,b,c) due to the different comparison cohorts.

The new CNV method results corroborate all the events highlighted by each of the existing method analyses as well as additional events which were not detectable or do not meet the thresholds of the existing Archer CNV method.

Performance

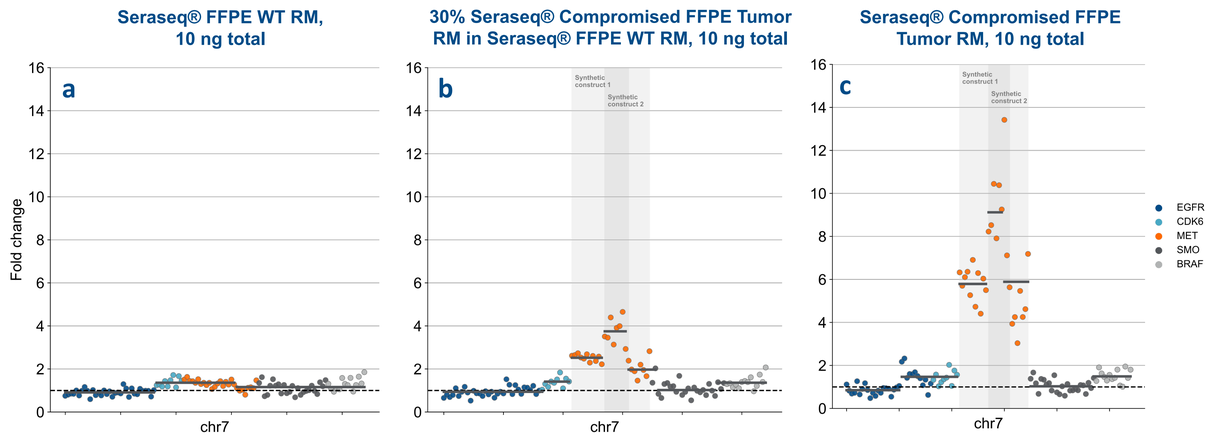


Figure 2. CNV breakpoint detection of synthetic MET amplification in challenging low input mass and contrived low aberrant cellularity libraries

VARIANT*Plex* libraries were prepared using 10 ng total input mass of either Seraseq FFPE WT DNA reference material (a), Seraseq Compromised FFPE Tumor DNA reference material (c), or a mix of 30% Tumor DNA reference material with the WT DNA reference material as background (b). Each point in the plot represents a primer in the panel, colored by the target gene, and dark gray lines are drawn at the mean fold change of segments. The black dashed line is the sample's baseline, fold change = 1, against which primer and segment fold changes were calculated. Results shown are representative of both 10 ng replicates of each input and 50 ng replicates of the same.

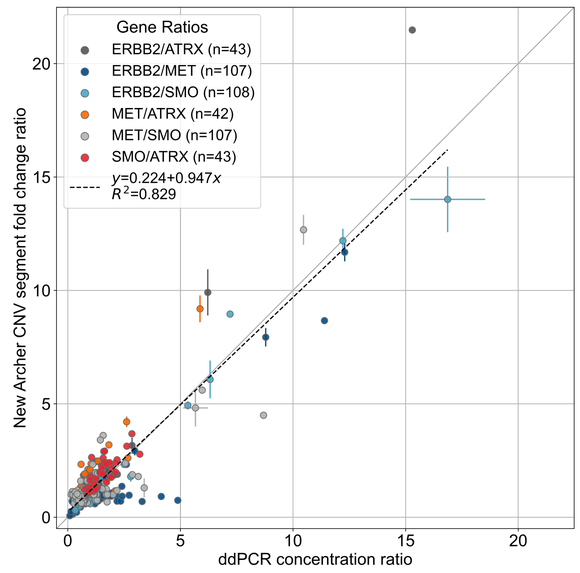
The manufacturer used two overlapping synthetic constructs to create the MET gene amplification in the Seraseq Compromised FFPE Tumor DNA reference material. The locations of these constructs are represented in the figures by regions highlighted in gray. A greater amplification is expected where these constructs overlap (darker gray region). Our CNV method was able to reliably and accurately detect the internal breakpoints that result from the overlap of these constructs, even in the 30% positive material mix library at 10 ng (b), without requiring normal tissue.

Figure 3. Concordance of CNV segment fold change ratios with ddPCR concentration ratios

The results of ddPCR and our new CNV method for 108 unique inputs, including tumor FFPE extracts, cell lines, and reference materials with a variety of amplifications and deletions are depicted in the figure. Points are colored according to the pair of genes compared. The x and y values are the ratios (or mean ratios when available) calculated from ddPCR concentrations or CNV segment fold changes for segments that overlap with the coordinates of the ddPCR probes used. Standard deviation error bars are drawn, and mean ratio plotted, when there were replicate libraries, replicate ddPCR results, or both.

Inputs were assayed with 4 ddPCR probes located in MET, ERBB2, SMO, and ATRX and with VARIANT*Plex* libraries prepared using panels ranging in size from approximately 900 to 10,000 primers. The number of points for each gene ratio pair is unequal because not all panels used to prepare libraries covered the regions that were targeted by ddPCR, e.g., all panels covered ERBB2 and SMO, but some panels did not also cover ATRX.

The agreement between our segment fold change ratios and the ddPCR concentration ratios indicates that our modelling and segmentation methods perform well across this diverse set of inputs.



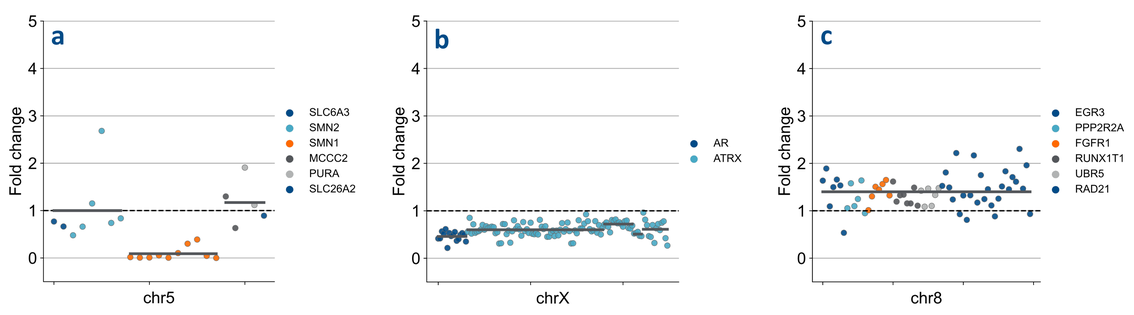


Figure 4. Selected representative CNV events detected with new Archer CNV method

The results shown here demonstrate the capability of our new CNV method to identify and call a variety of CNV events in different applications. In each plot, points are primers colored by the gene they target. The dashed black line is the baseline copy number, fold change = 1. Dark gray lines represent the mean fold change for a segment.

In (a), a homozygous deletion of SMN2 was detected in a library prepared with 50ng germline input with a small (ca. 500 primers) custom VARIANT*Plex* panel. We frequently observe heterozygous 'losses' of X and Y in XY samples, as in (b). We have observed whole chromosome gains as in (c) which depicts a case of trisomy 8. In addition to being able to detect the above one or two copy number events, some of which span whole chromosomes, (d) depicts very large copy number gains detected in MYCN and exons 3 & 4 of ALK in a neuroblastoma cell line.

Conclusions

The new Archer CNV method, which does not use any paired or panel of normals, performs well on diverse input types, across a range of panel sizes, and can detect CNV events that vary in genomic span and fold change magnitude. As with any method, the specific size (in bp) resolution is dependent on the distribution of probes used - primers in the panel, in our case - and events of smaller copy number difference may be limited by input purity. Results shown here demonstrate the agreement of our fold change measurements with ddPCR (fig. 3), the ability of this method to accurately detect the breakpoints of sufficiently large gains in as little as 30% aberrant input (fig. 2b), and the ability to detect CNV events as large as whole chromosomes (fig. 4b,c) or as small as two exons (fig. 4d).

Performing CNV detection using our new method allows reliable and accurate detection of CNVs without the overhead or confounding factors that comparisons to normals or a panel of normals can introduce.

Taylor R. Patterson
Assay Development R&D, Archer (Integrated DNA Technologies)
tpatterson@idtdna.com