Tumor Only Allele Specific Copy Number Calling with a Targeted Next-Generation Sequencing Panel

Allison Hadjis, Taylor R. Patterson, Mark Rogers, Laura Johnson, Trent Fridey, David McConnell | Integrated DNA Technologies

I037

Introduction

Allele specific copy number (ASCN) algorithms expand the options for relevant insights gained from next-generation sequencing (NGS) data by providing the number of copies of the major and minor alleles in a targeted region. Further, patterns in ASCN can indicate genomic instability resulting from homologous recombination deficiency (HRD). ASCN has benefits over conventional copy number variation (CNV) because it allows identification of regions with copy neutral loss of heterozygosity (LOH), see Figure 1. Additionally, ASCN offers deeper insights into copy number variation by allowing tumor purity and ploidy estimation not permitted by CNV alone. Many ASCN algorithms require a matched normal, use an internal panel of normals, or make assumptions on tumor purity and/or ploidy to estimate the allele copy numbers, which introduces error. Moreover, many ASCN algorithms rely on whole exome or whole genome sequencing to get sufficient data, which requires 4-50x more read depth than Archer's targeted panel which only requires 12M reads. Here, we present a tumor only ASCN algorithm capable of estimating tumor purity and ploidy using a targeted NGS panel that is compatible with Archer Anchored Multiplex PCR (AMP™) chemistry and VARIANT*Plex*™ panels.

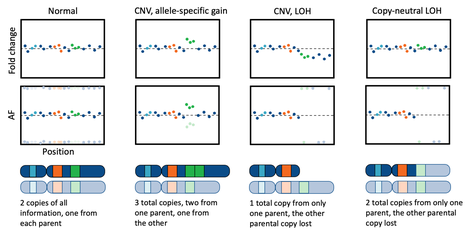


Figure 1. Graphic depicting CNV and Allele Frequency (AF) data in different scenarios.
The top row shows the CNV, or fold change, data for a region of the genome, the middle row shows the AF distribution, and the bottom row shows the example parental chromosomes for each scenario. Follow the "green gene" to see normal, CNV allele specific gain, CNV LOH, and copy-neutral LOH situations. Incorporating ASCN information helps identify copy-neutral LOH situations shown in the last column where CNV is blind to changes, but the AF plot reveals an LOH.

Methods

This tumor only ASCN algorithm relies on a panel targeting 5000 single nucleotide polymorphisms (SNPs) evenly spaced throughout the genome that are likely to be heterozygous, see Table 1 for panel details. CNV and SNP calling on reads from AMP libraries is performed by Archer Analysis. The SNPs are filtered to remove likely homozygous variants and then segmented. Tumor purity is estimated using somatic variants in targeted genes and tumor ploidy is calculated using a decision tree followed by a grid search. The total and minor copy number for each segment is then calculated. The algorithm was tested on two tumor/normal cell line dilution series (one with 11 dilution points from 10%-90% tumor and one with 7 dilution points from 10%-50% tumor), four cell lines of known ploidy, five cell lines with known ASCN breakpoints, and samples with orthogonal comparison from Repare Therapeutics.

	Reads	Genes
VARIANT*Plex* Comprehensive Tissue and Blood	61 M	460
VARIANT*Plex* Complete Solid Tumor v2	57 M	430
VARIANT*Plex* Pan Solid Tumor v2	37 M	185
VARIANT*Plex* HRD Module *	12 M	-

* HRD module contains 5062 SNP Array primers and 131 HRD primers

Table 1. SNP Array panel offerings.
The SNP Array comes in a standalone module, VARIANT*Plex* HRD Module, with 5062 SNP Array primers and 131 HRD specific primers. This module requires 12M reads. The SNP Array is also included by default in the VARIANT*Plex* Pan Solid Tumor v2, the VARIANT*Plex* Complete Solid Tumor v2, and the VARIANT*Plex* Comprehensive Tissue and Blood panels.

Results

In five cell lines with known breakpoints and regions of LOH, Archer ASCN has an accuracy of 72% across 43 events[2] (Table 2 and Figure 2). Additionally, the ASCN can accurately identify abnormal ploidy in four different cell lines CRL-2338D (tetraploid), CRL-5868 (near triploid), 786-0 (hypertriploid), and K-562 (triploid) (Table 3 and Figure 3). The estimated purity and the expected purity of the tumor/normal cell line dilution series have a high concordance with an R^2 of 0.93 and can detect tumors as low as 20% tumor purity (Figures 4 and 5). Lastly, Archer ASCN is concordant with an orthogonal method (Figures 6, 7, 8 and Table 4).

Breakpoint, Ploidy, and Purity Results

High accuracy on cell lines with known breakpoints

Cell Line	# Total Expected Events	# Archer Called Correctly	Accuracy
IMR32	8	7	88%
SK-N-BE(2)	12	5	42%
SK-N-DZ	8	6	75%
SK-N-FI	9	8	89%
SK-N-SH	6	5	83%
Total	*43*	*31*	*72%*

Table 2. Breakpoint accuracy of cell lines.
Five cell lines with known gain, loss, and LOH events[2] were ordered from ATCC. 50ng of input was prepared with the VARIANT*Plex* Pan Solid Tumor v2 panel and sequenced on the NextSeq 550 to a read depth of at least 30M reads. Analysis was performed on Archer Analysis 7.4. Overall, there were 43 expected gain, loss, and LOH events[2]. Archer ASCN called 31 correctly for an accuracy of 72%. Detailed results for individual sample SK-N-F1 are shown in Figure 2.

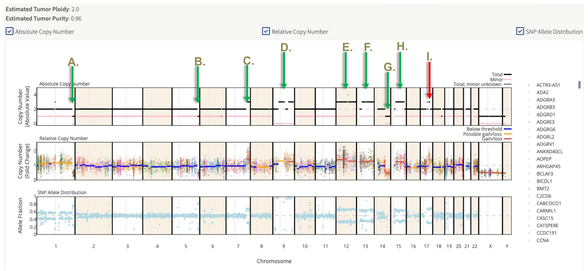


Figure 2. Cell line SK-N-F1 breakpoint results.
Cell line SK-N-F1 is a known breakpoint sample from Table 2.This screenshot comes directly from the Archer Analysis 7.4 UI with green arrows indicating correctly called events and a red arrow indicating an incorrectly called event. The UI is interactive allowing for zooming and gene selection. There are 9 expected events in this cell line and Archer ASCN called 8 correctly for an accuracy of 89%. Events called correctly are chromosome 1 telomere loss (A), chromosome 5 telomere loss (B), chromosome 7 telomere gain (C), chromosome 9 LOH (D), chromosome 12 gain (E), chromosome 13 gain (F), chromosome 14 q arm loss (G), and chromosome 15 gain (H). The only event missed in this sample is chromosome 17 p arm LOH (I).

Accurate ploidy estimates demonstrated in aneuploid cell lines

Cell Line	Expected Ploidy from ATCC	Archer ASCN Ploidy
CRL-5868	"near triploid"	3.0
CRL-2338D	"tetraploid"	4.1
K-562	"triploid"	2.9
786-0	"hypertriploid"	3.5

Table 3. Ploidy estimation performance on cell lines with known ploidy.
Archer ASCN has accurate ploidy estimation on cell lines with known ploidy. Four cell lines with known ploidy were ordered from ATCC. The samples were prepared with 50 ng of input and sequenced on a NextSeq 550. Analysis was performed on Archer Analysis 7.4. Cell lines CRL-5868 and CRL-2337D were prepared using the VARIANT*Plex* Pan Solid Tumor v2 panel and sequenced to a read depth of at least 30M reads. Cell lines K-562 and 786-0 were prepared using the VARIANT*Plex* Complete Solid Tumor v2 panel and sequenced to a read depth of at least 60M reads. Detailed results for individual sample K-562 are shown in Figure 3.

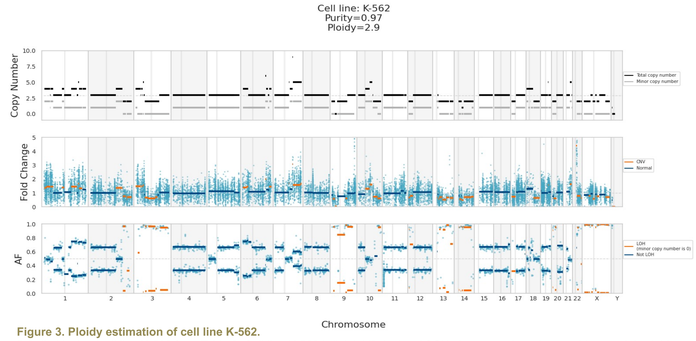


Figure 3. Ploidy estimation of cell line K-562.
Detailed look at cell line K-562 from Table 3. Cell line K-562 is "triploid". Archer ASCN also correctly estimates this sample's ploidy at 2.9 and there is strong evidence of this call in the AF distribution seen in the 33% and 66% AF split.

Estimated purities correlate with expected purities

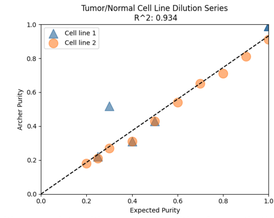


Figure 4. Archer purities correlate with expected purities for synthetic dilution series.
Overall correlation between the expected purity and Archer ASCN purity has an R^2 of 0.93 and a minimum purity estimate of 20%, see Figure 5 for details of this sample. Two tumor cell lines were diluted in their normal counterparts to create two distinct tumor normal dilution series. Cell line 1 is made of tumor, CRL-5868, and matched normal, CRL-5957 (BL). Cell line 1 has dilutions from 10%-50%. Cell line 2 is made of tumor, CRL-2338D, and matched normal, CRL-2339D (BL). Cell line 2 has dilutions from 10%-90%. 50 ng of both cell line dilution series were prepared with the VARIANT*Plex* Pan Solid Tumor v2 panel and sequenced on a NextSeq 550 to a minimum read depth of 27M reads. Analysis was performed on Archer Analysis 7.4.

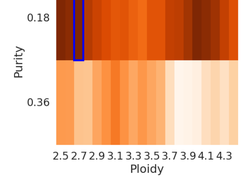


Figure 5. Grid search for best purity and ploidy solution for Cell line 2, 20% tumor purity sample.
The Archer ASCN purity and ploidy algorithm correctly identifies this sample, which is the 20% tumor from Cell line 2 in Figure 4. Using information from somatic variants in targeted genes (both heterozygous and homozygous) and ploidy options determined from the AF profile, a grid search is performed to find the optimal purity and ploidy fit across all the segments.

Orthogonal Method Validation

High concordance of breakpoints and LOH calls

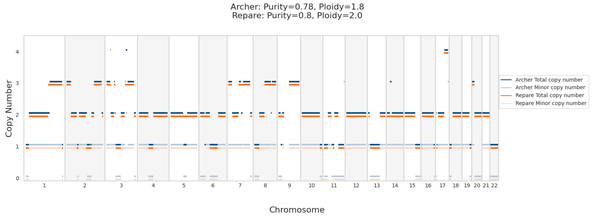


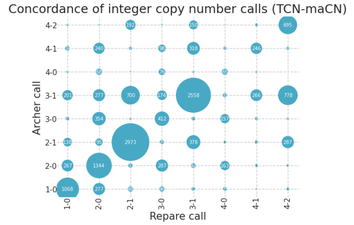
Figure 6. Archer and Repare total and minor copy number estimates for single sample.
99% of total and minor copy number calls overlap between Archer ASCN and an orthogonal method. This sample is prostate tissue with biallelic BRCA2 mutations. The sample was prepared with VARIANT*Plex* Repare v2 panel with the SNP Array. The sample was analyzed in Archer Analysis 7.4 and with orthogonal method, Repare[1], which uses an extensive and highly curated panel of normals and FACETS[3].

		Archer	
		LOH	Non-LOH
Repare	LOH	4,635	1,360
	Non-LOH	248	9,605

Table 4. LOH calling confusion matrix.
LOH events are called similarly between Archer Analysis 7.4 ASCN and orthogonal method, Repare[1] , which uses an extensive and highly curated panel of normals and FACETS[3]. LOH overall percent agreement is 90%, LOH positive percent agreement is 95%, and LOH negative percent agreement is 88%. In this comparison 292 samples with percentage duplicate reads <50%, mean library insert size >100, mean coverage >1000, and purity in both methods > 30% were used. Segments were filtered to be >5Mb and >25% overlap.

Integer copy numbers agree

Figure 7. Bubble plot comparison of integer copy number calls for Archer ASCN and orthogonal method.
Archer ASCN and Repare[1] show high concordance of integer copy number calls with 59% of Total Copy Number – Minor Allele Copy Number (TCN-maCN) intersected segments falling along the diagonal which indicates the same call was made for a segment. In this comparison 292 samples with percentage duplicate reads <50%, mean library insert size >100, mean coverage >1000, and purity in both methods > 30% were used. Segments were filtered to be >5Mb and >25% overlap.



Archer ASCN purities correlate with orthogonal method

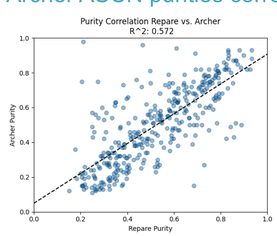


Figure 8. Archer purities correlate with orthogonal method purities.
Correlation between orthogonal method, Repare[1], and Archer Analysis 7.4 ASCN purity has an R^2 of 0.57. In this comparison 373 samples with percentage duplicate reads <50%, mean library insert size >100, mean coverage >1000, and at least one somatic variant in a targeted gene were used. All samples were prepared using the VARIANT*Plex* Repare v2 panel with the SNP array. The orthogonal method uses an extensive and highly curated panel of normals and FACETS[3].

Conclusions & Acknowledgements

Tumor only ASCN with a targeted NGS panel and AMP chemistry can estimate tumor purity and ploidy leading to accurate identification of total and minor allele copy numbers. Archer ASCN has demonstrated high concordance and accuracy with expected results. Patterns from ASCN results can inform HRD status. For more information on Archer HRD calling, please see posters ST140 and ST105.

The authors gratefully acknowledge the helpful discussions and samples provided by Ian Silverman and Repare Therapeutics.

References

1. Glodzik, D., Selenica, P., Rogge, R. A., Silverman, I. M., Mandelker, D., Harris, S., Zhao, J., Zinda, M., Veloso, A., Malani, N., Riaz, N., Koehler, M., Daber, R. D., Johnson, V., Rimkunas, V., & Reis-Filho, J. S. (2023). Detection of Biallelic Loss of DNA Repair Genes in Formalin-Fixed, Paraffin-Embedded Tumor Samples Using a Novel Tumor-Only Sequencing Panel. The Journal of molecular diagnostics : JMD, 25(5), 295–310. https://doi.org/10.1016/j.jmoldx.2023.02.004
2. Kryh, H., Carén, H., Erichsen, J. et al. Comprehensive SNP array study of frequently used neuroblastoma cell lines; copy neutral loss of heterozygosity is common in the cell lines but uncommon in primary tumors. BMC Genomics 12, 443 (2011). https://doi.org/10.1186/1471-2164-12-443
3. Shen, R., & Seshan, V. E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic acids research, 44(16), e131. https://doi.org/10.1093/nar/gkw520

Allison Hadjis
Bioinformatics Research Scientist,
Archer (Integrated DNA Technologies)
ahadjis@idtdna.com