AccuGenomics
A higher standard of accuracy

SNAQ™-SEQ ThermoFisher OCAv3 CNV Detection at Exon Level, Normal and Tumor Samples

Tom Morrison, AccuGenomics, Inc.

Bryan Lo, The Ottawa Hospital

Venus Chirip, The Ottawa Hospital

2024-02-16

Abstract

The loss of tumor suppressor gene copies is an important cancer driver. However, detecting genomic copy loss by NGS in cancer FFPE specimens is not very robust. In this white paper we examine the ability of SNAQ™-SEQ internal standards (IS) to improve CNV detection when integrated into the Thermofisher Oncomine Comprehensive Assay v3. Internal standards were designed to biochemically covary with ATM, BRCA1, BRCA2 & PALB2 exonic regions, allowing precise genomic abundance determination for each exon, from which exon copy loss would be detected. SNAQ™-SEQ gave exon level copy results for all six normal tissue and all nine tumor samples, whereas the Thermofisher OCAv3 pipeline reported only half of these genes, and their values are limited to whole gene level. Four of the missing OCAv3 coincided with gene copy loss. Overall, SNAQ™-SEQ CNV detection was more robust and accurate than the existing OCAv3 method.

Introduction

Accurate copy number variant (CNV) detection by NGS needs to account for sources of testing bias. NGS, whether by hybrid capture, PCR primer enrichment, or direct long read sequencing, produces sequencing reads disproportional to their starting levels. Further, because of the high complexity of NGS testing, small drifts in reagents and instruments reduce reliability of CNV methods that depend on external calibration samples. Further, use of UMI/molecular tags to count unique templates does not fix this issue as end repair/ligation bias and deduplication errors distort the final read level of input templates. What is needed is a method to detect CNV that eliminates the testing bias that has long been responsible for unreliable CNV detection.

Since the days of qPCR, internal standards have been recognized as a method to correct for testing bias (e.g., competitive PCR, or analytic mass spectrophotometry). Over the last ten years, AccuGenomics has developed Standardized Nucleic Acid Quantification (SNAQ™-SEQ) IS that biochemically mimic the native template behavior during library preparation, sequence detection and bioinformatic analysis. As a result, the internal standards covary with the native template and allow for within sample quality controls such as variant sensitivity, standardized abundance, complexity capture and limit of blank. For CNV detection, SNAQ™-SEQ IS enable a standardized abundance measurement with similar

accuracy as droplet digital PCR but with the advantage of multiplexed NGS throughput. Once the IS is added to a sample, the molar ratio between IS and native templates (NT) is maintained throughout the testing process, and ultimately IS levels quantified in the VCF through detection of the unique IS variant. Like the method developed for competitive PCR, by knowing the quantity of IS added to sample and the resulting NT:IS ratio, the NT abundance at the IS variant position may be calculated. In this report, we demonstrate the use of Standardized Nucleic Acid for Sequencing (SNAQ™-SEQ) Internal Standards (IS) designed to biochemically mimic the native exon templates of ATM, BRCA1, BRCA2 and PALB2. CNV detection was performed on FFPE samples by integrating SNAQ™-SEQ into the ThermoFisher OCAv3 NGS panel. The CNV reporting rate, false negative, and false positive rates will be compared between samples run with or without the SNAQ™-SEQ IS.

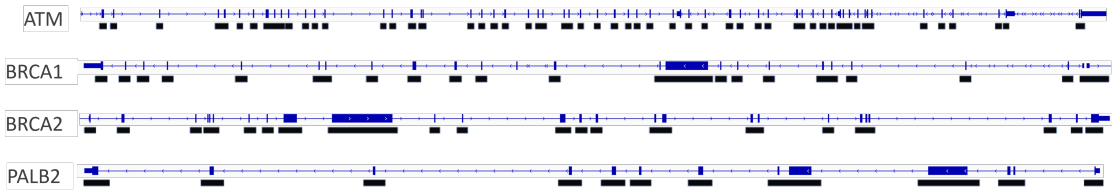


Figure 1. SNAQ™-SEQ Mammo DNA Internal Standards. The figure depicts the human genome mapping of exon locations (NT) and corresponding internal standards (IS) for ATM, BRCA2, BRCA1 and PALB2 gene regions.

Method

Accukit™ Mammo DNA IS were designed to each exon of the ATM, BRCA1, BRCA2 and PALB2 genes (see Figure 1). Each exonic amplicon had at least two unique IS SNP for abundance measurement. Purified DNA (20 ng) from nine FFPE sections ranging from 40% to 90% tumor cellularity previously evaluated by ThermoFisher OCAv3, or six expected normal copy level sections were spiked with 200 copies IS prior to running the ThermoFisher OCAv3 test. VCF created using the IonReporter pipeline utilized a modified hotspot.bed file to ensure the IS variants were reported. An R markdown script extracted the IS variant counts from the VCF, calculated the exon abundance from NT:IS read count ratio, and the abundance then used to detect gene copy loss. The following QC was performed on each exon abundance measurement:

- At least two exonic IS variants had a PASS FILTER from the Ion Reporter,
- IS raw alt count > 50,
- IS VAF was >1%,
- the exon IS variants VAF must be within 10% of each other.

The last two requirements were implemented to address IonReporter variant calling errors that arose from end repair artifacts. Insufficient IS added into sample was flagged based on the premise that 200 copies added to 5500 FFPE damaged haploid genomes (20 ng) should produce an IS VAF about 3-fold above 1.3%. Low IS input can lead to insufficient exon measurements for CNV calculation. Too little NT input, flagged if the NT abundance was <300 copies, could indicate a sample preparation issue and give rise to imprecise abundance measurement. Next, the gene abundance was converted to copies per gene. To

calculate the abundance that corresponded to “normal” 2 gene copies per cell, a modified-Z score was used to eliminate abundance outliers prior to calculating the median abundance of an equal number of randomly selected exons for each gene (usually 13 exons per gene). The median exon abundance was used to scale abundance to copies per cell. A gene copy loss was indicated when the combined exon copy number 95% standard error confidence interval was less than 1.8.

The SNAQ™-SEQ copy loss detection was compared with the original OCAv3 test results which calculates a similar standard error copy number interval.

Results

The plots in APPENDIX 1 depict the SNAQ™-SEQ copy number analysis of each FFPE patient sample (right plots), indicating genomic abundance (left y-axis) or copies per cell (right y-axis) for each exon (circles) colored by gene name (x-axis). The x-axis region indicated by “LOAD” were single exons from GENE/EXON: CDK12/09, CHEK1/02, NBN/02,15 PIK3R1/02 RAD41/05 and will not be discussed in this white paper. Each exon abundance calculation failing SNAQ™-SEQ QC were indicated by ‘X’ at bottom of each plot. The red lines indicate a modified Z-score 3-sigma cutoff used to remove outliers prior to establishing “normal” abundance.

Table 2 depicts a SNAQ™-SEQ CNV report for these samples. Two samples, F & R indicated possible low levels of IS added to sample, and consequently the reduction in exon abundance measurements likely arose from sub 1% VAF, and the chromosome abundance should be overestimated because it assumes 200 IS copies per sample. In the case of sample R, the IS variants were mostly too low to be detected by OCAv3 which triggered rerunning the sample (O) with the correct amount of IS input. No samples had an abnormally low level of NT abundance. None of the “normal” samples (C - G, Q) indicated copy loss. When copy loss was detected, we suggest referring to the plots to visually confirm the level of copy loss. For example, sample H and its repeat sample K have BRCA1, BRAC2 single gene loss and an additional partial BRCA1 loss.

Table 1 compares the SNAQ™-Seq results with OCAv3 original analysis. OCAv3 copy analysis often had insufficient statistics to give a PASSing CNV analysis. Of the nine overlapping samples, OCAv3 did not give CNV results for half the genes (18 of 36 genes), four of which were indicated as copy loss by SNAQ™-SEQ analysis (I-BRCA2, H-BRCA1, L-BRCA1, O-BRCA1), suggesting OCAv3 testing may miss important copy loss detection. Further, the simple OCAv3 approach CNV would unlikely detect the partial BRCA1 double copy loss of sample H. Lastly, two OCAv3 copy loss estimations of samples A and J were not supported by SNAQ™-SEQ abundance, suggesting OCAv3 method may have indicated false positive.

Table 1. Summary of Copy Loss Detected by SNAQ™-SEQ or OCAv3 pipeline.

SAMPLE	SNAQ™-SEQ	OCAv3
A	BRCA2	(ATM) BRCA1* BRCA2
B	BRCA1 BRCA2	(ATM BRCA1 BRCA2 PALB2)
H[2]	BRCA1[p] BRCA2 PALB2	(BRCA1) BRCA2 PALB2
I	BRCA2	(ATM BRCA1 BRCA2 PALB2)
J	BRCA2	ATM* (BRACA1) BRCA2
K[2]	BRCA1[p] BRCA2	ND
L	BRCA1 PALB2	(BRCA1) PALB2
M	none	(BRCA1 BRCA2 PALB2)
N	none	(BRCA1)
O[1]	BRCA1	(BRCA1 BRCA2)
P	none	ND

KEY
SAMPLE letters correspond to plots in Appendix A
none No copy observed copy loss
ND No Data
***** SNAQ™-SEQ results do not support copy loss
GENE indicates loss of copy detected
(GENE) CNV analysis of gene did not pass QC.
[p] partial gene copy number loss
[1] Sample R had insufficient IS added, Sample O was a repeat
[2] Sample H was a repeat of sample K

Additional Observations

Each sample had differing levels of noise as indicated by the differences of the 3 x MAD boundary separation between samples. For normal DNA samples, the amount of genome captured as sequencing reads varies by primer efficiency, molecular tag addition, clonal amplification, etc. However, FFPE samples have the additional complication of differing regional genomic DNA damage due to fixation events such as DNA fragmentation, cross linking to cellular matrix, genomic structures (e.g., nucleosomes, chromatin, DNA binding proteins), and base alterations, varying by sample and genomic region. As the IS templates were undamaged, they were unable to mimic the pre damage associated with FFPE treatment. As a result, SNAQ™-SEQ will eliminate the testing bias, but the sample's FFPE damage will complicate detection of altered gene copy. For example, samples N or Q (see plots in Appendix A) had remarkably similar inter exon abundances suggesting uniform genomic damage from which it is readily apparent there was no altered gene copy. However, sample B and O copy levels were less conclusive; the higher ATM levels could arise from: less ATM genomic fixation damage; or/and the other genes were more affected

by fixation; or/and the ATM gene had genomic amplification; or/and single copy loss of BRCA1, BRCA2 & PALB2.

Samples E, G & H demonstrated a significantly low abundance for the first exon in PALB2. Further investigation is required to determine if low copy number arose from a genomic copy alteration, or localized FFPE damage. One cause for a single lower exon abundance would be a SNP overlapping the primer binding sites that results in a reduced amplification efficiency.

As discussed above, the UMI/molecular-tag-based chemistry used to estimate copy number should distort the original genomic levels based on local sequence/sample damage differences and loss of complexity during deduplication. The plot below correlates SNAQ™-SEQ abundance when calculated using the deduplicated reads (FAO & FDP) vs raw reads (AO & DP). Greater than 70% of the deduplicated read exon abundance measurements were noisier. Based on these results, we suggest using raw alt counts when possible for SNAQ™-SEQ abundance calculations.

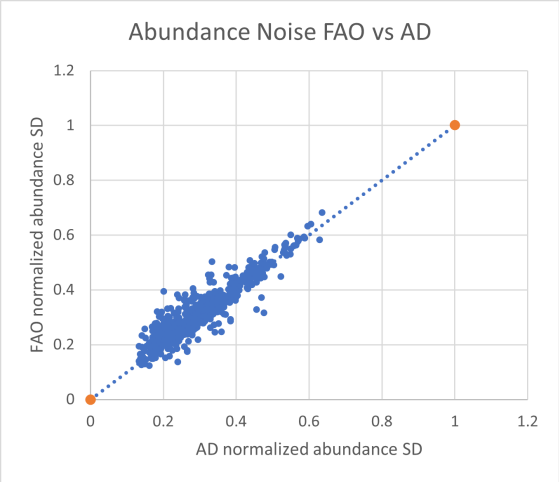


*Figure 2. Comparison of SNAQ™-SEQ abundance using either FAO or AD for NT & IS read counts. The separation of exon abundance from center (MAD*1.46 method) when using FAO (the IonReporter deduplicated read count, y-axis), or AD (the raw read count, x-axis) to calculate abundance. Higher values indicate noisier sample abundance measurement*

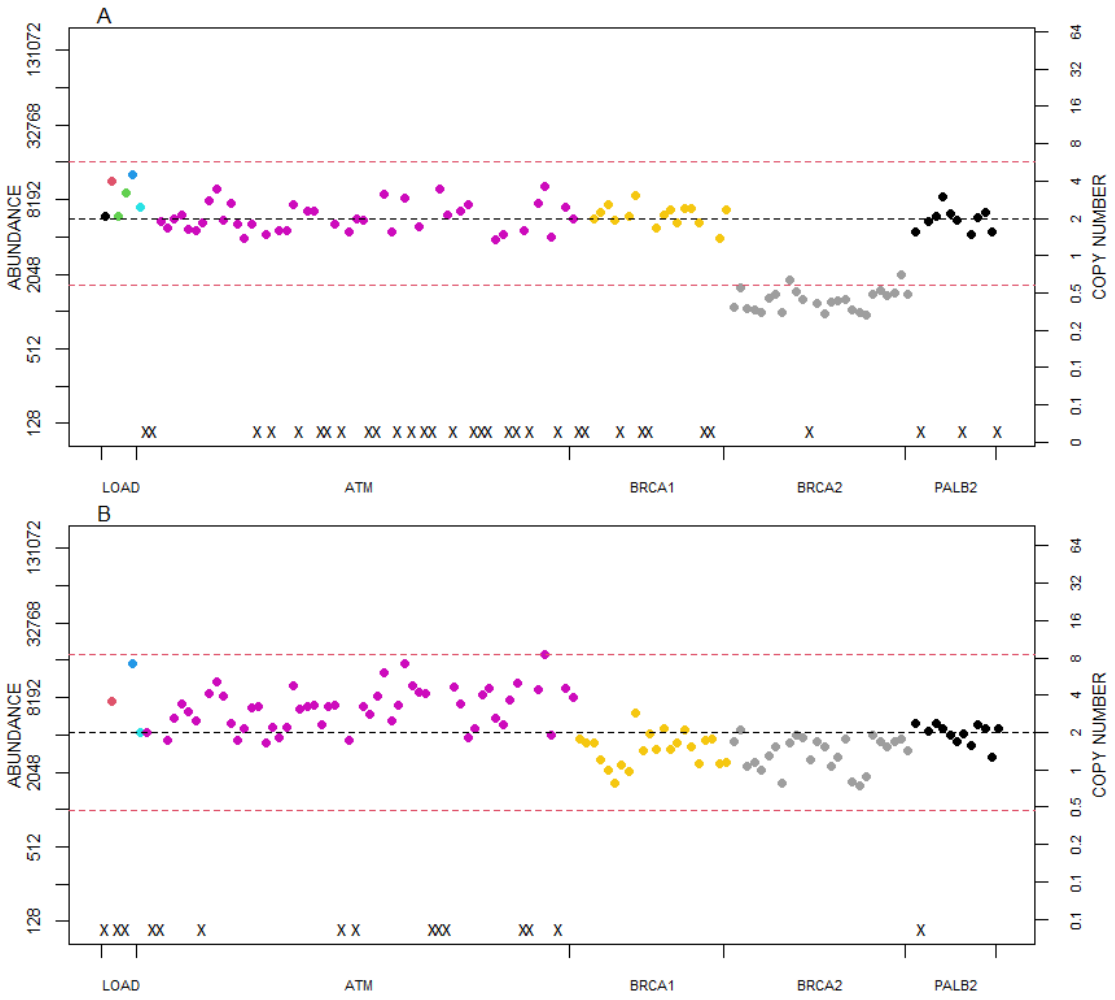
Conclusions

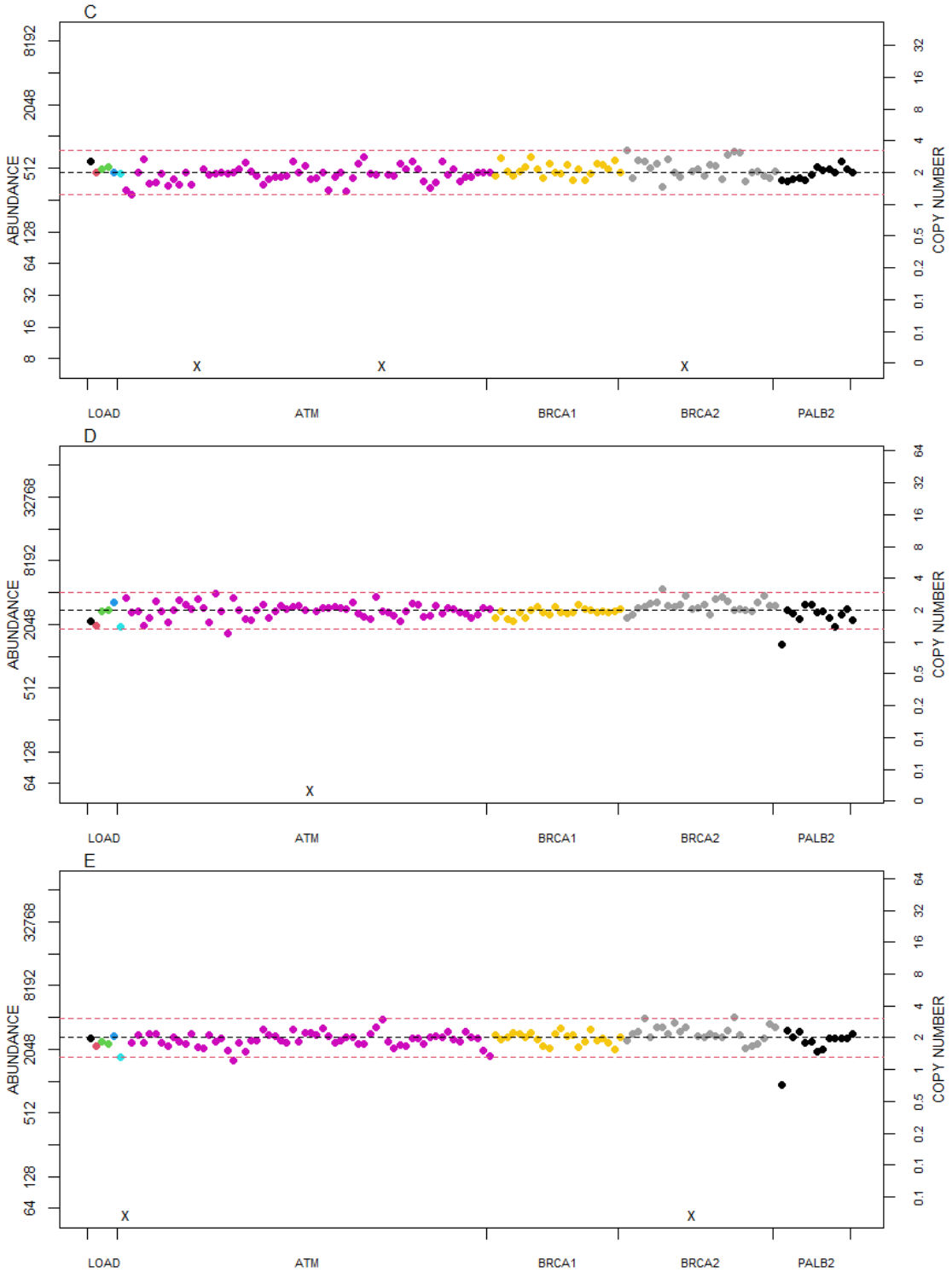
SNAQ™-SEQ copy loss detection was based on estimating the exon abundance of ATM, BRCA1, BRCA2, PALB2 genes, converting abundance into copies per cell and then detecting genes with significantly lower copy number. PASS/FAIL criteria based on IS input level or NT abundance, as well as copy loss detection proposed in this paper, should be adjusted to fit a lab's experience during validation. The CNV detection comparison between OCAv3 and SNAQ indicated SNAQ was more robust, giving a result for every gene in every sample, unlike OCAv3 reporting 50% of targets that resulting in missing copy loss mutations. Further, there was OCAv3 2 copy loss not supported by SNAQ™-SEQ analysis which could be OCAv3 false positives. Overall, SNAQ™-SEQ CNV detection routinely outperformed the

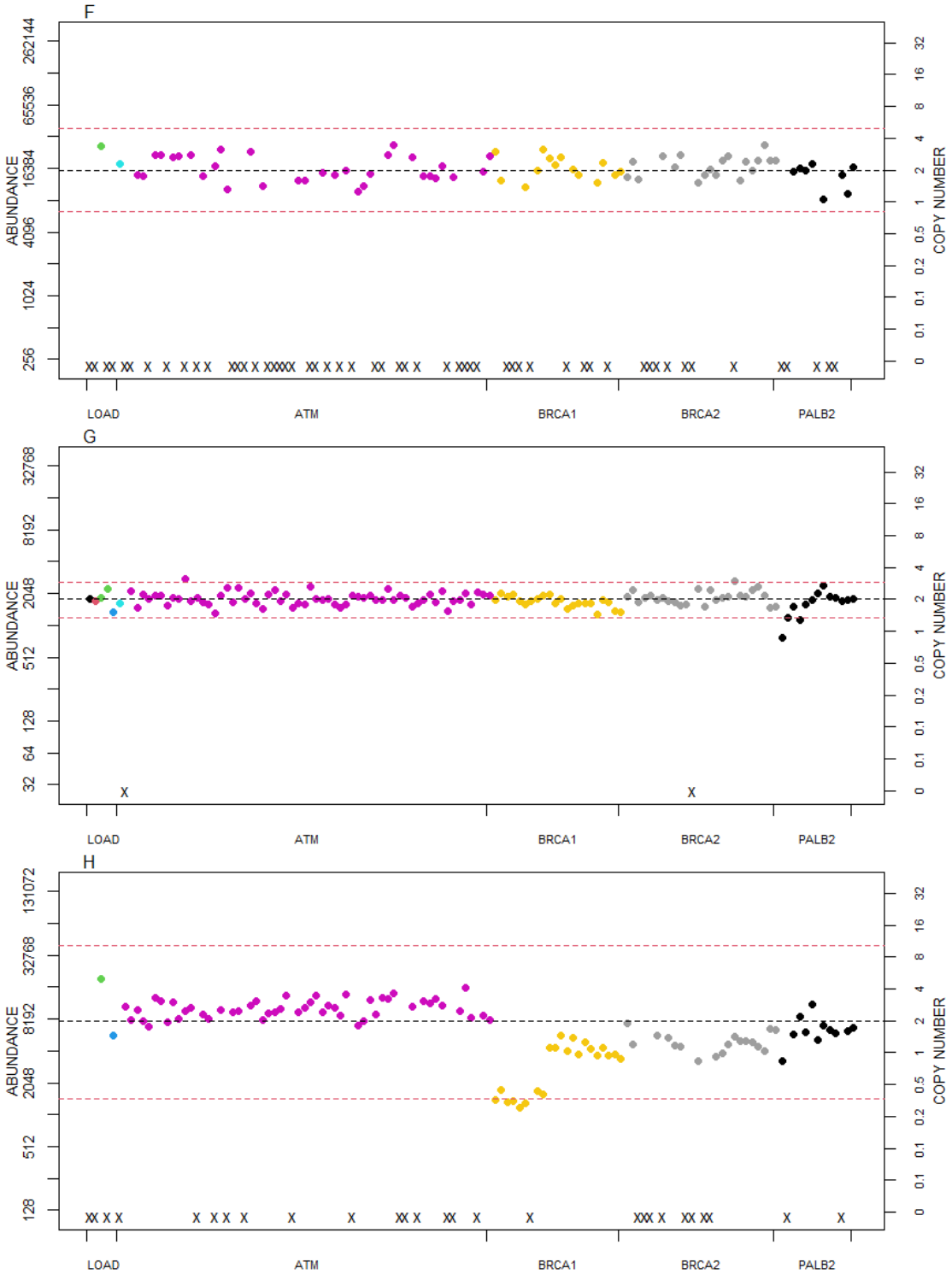
OCAv3 pipeline, warranting moving forward with its integration into the OCAv3 for testing of FFPE CNV.

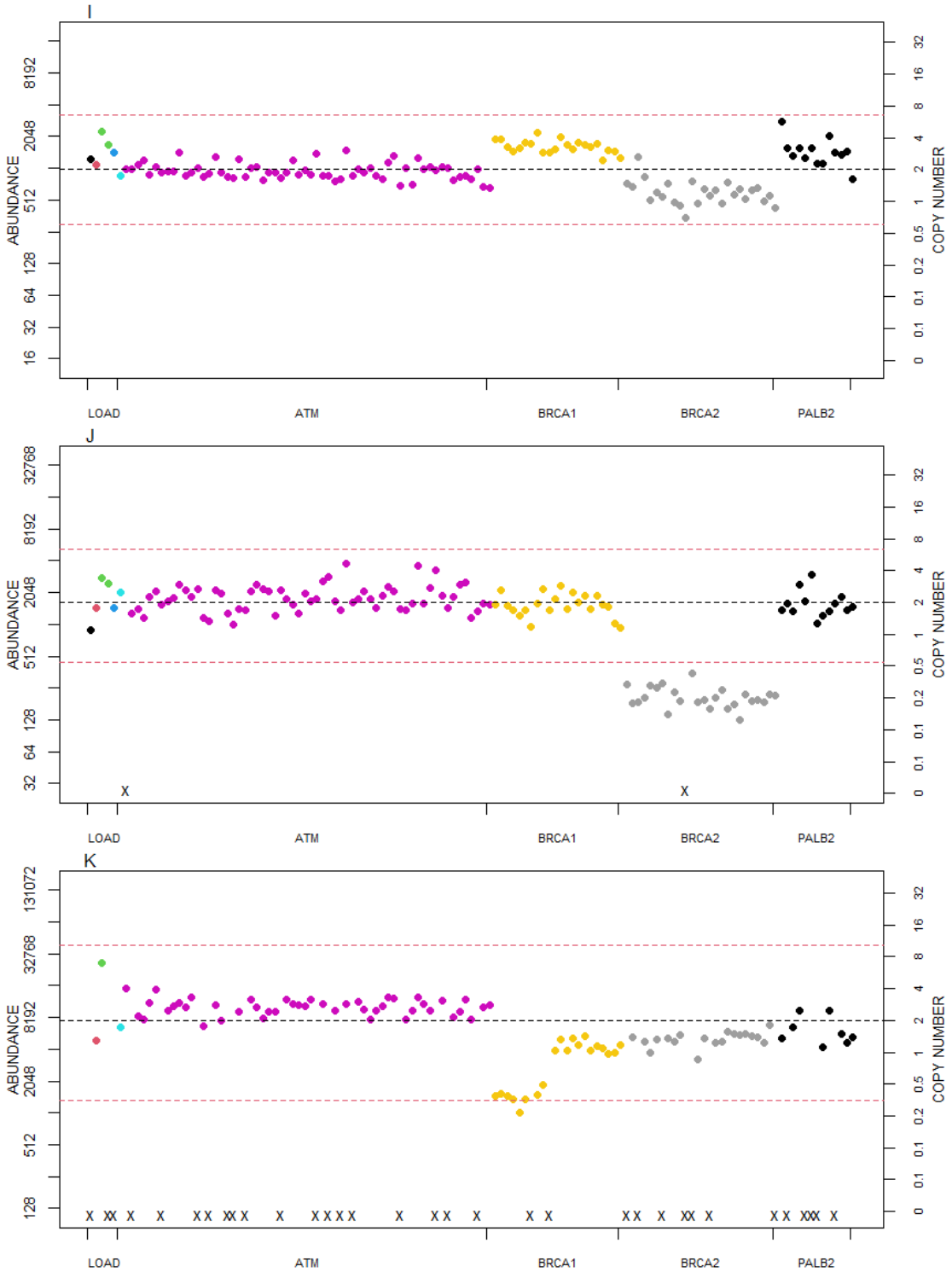
APPENDIX 1

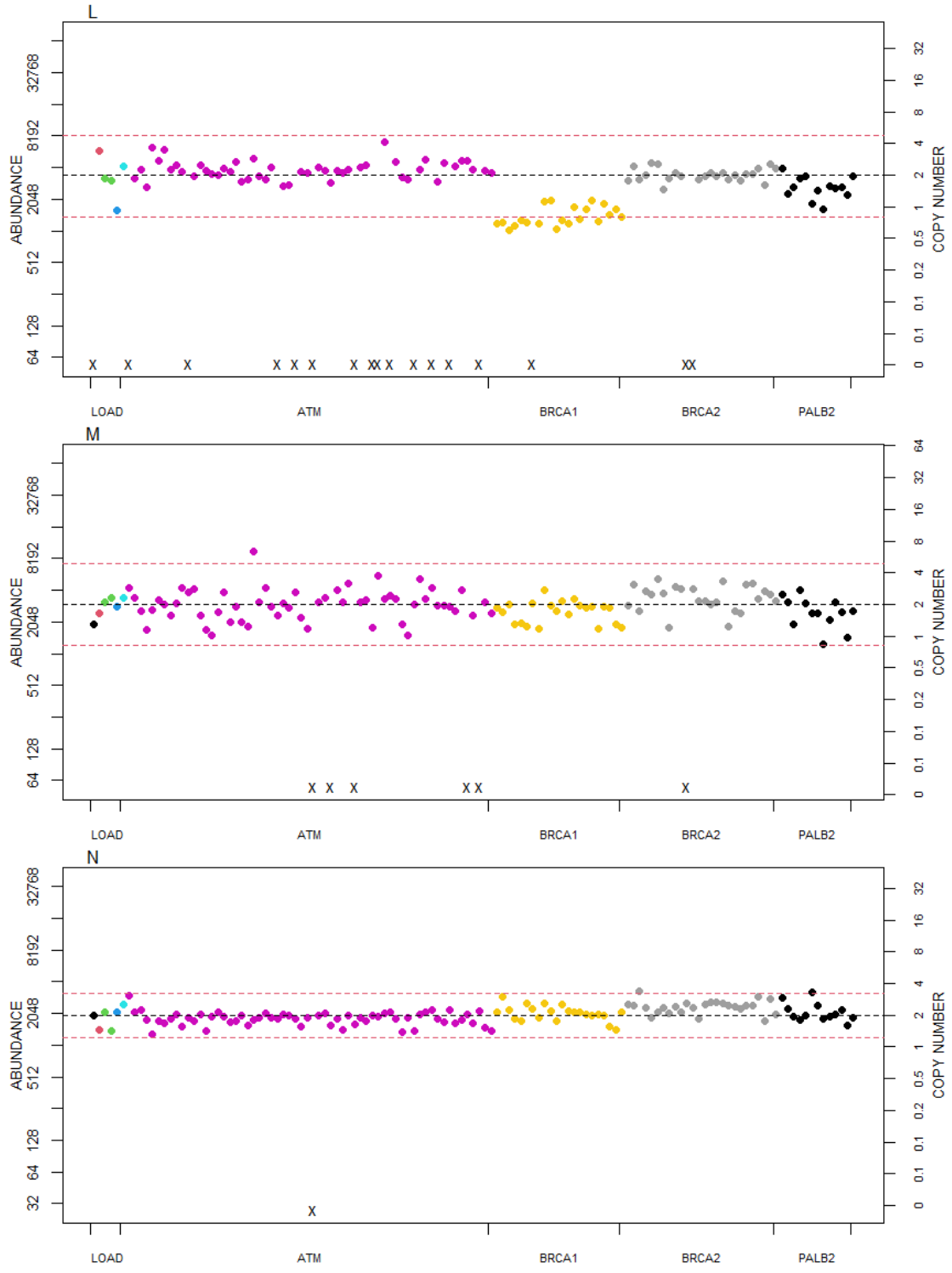
SNAQ™-SEQ CNV analysis plots.

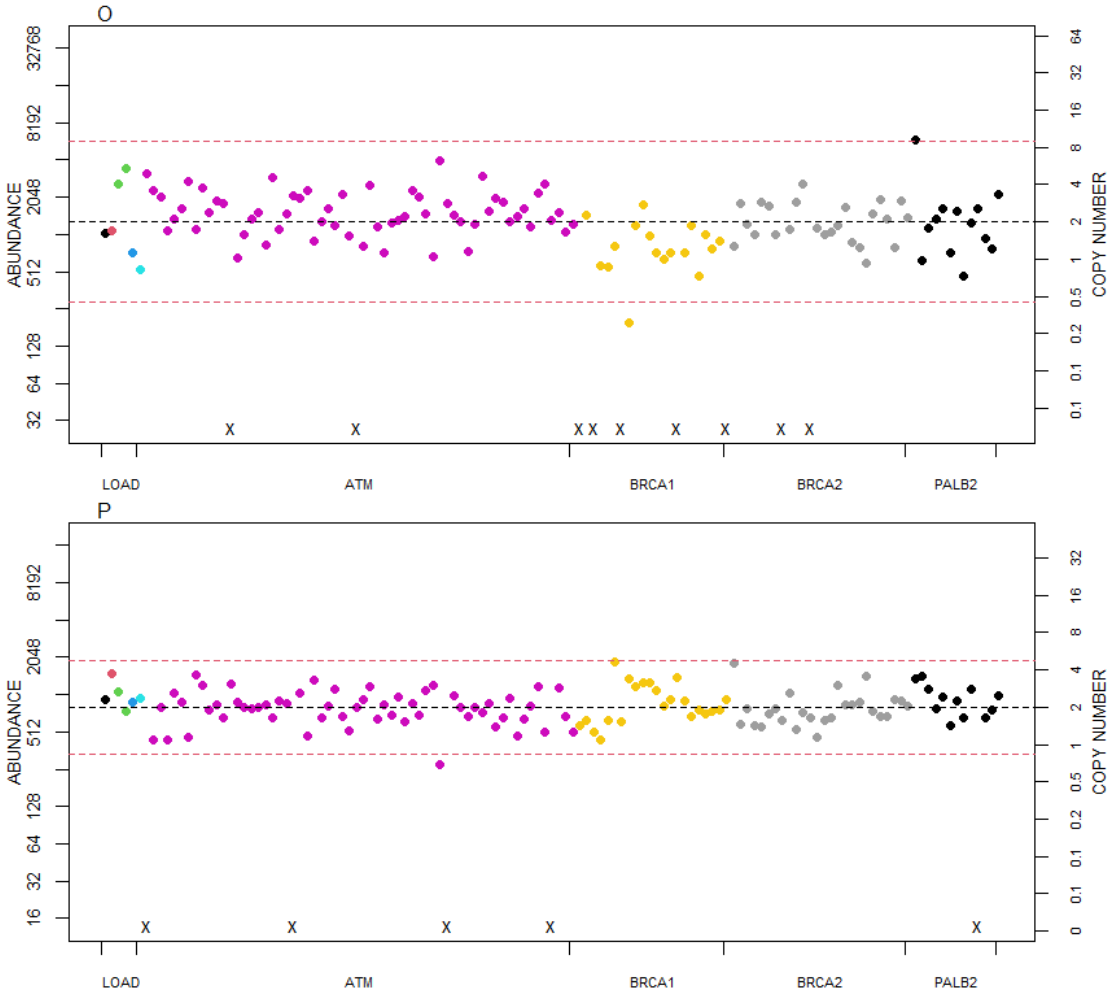



C
ABUNDANCE
8192
2048
512
128
64
32
16
8
COPY NUMBER
32
16
8
4
2
1
0.5
0.2
0.1
0.1
0
X
X
X
LOAD
ATM
BRCA1
BRCA2
PALB2
D
ABUNDANCE
32768
8192
2048
512
128
64
COPY NUMBER
64
32
16
8
4
2
1
0.5
0.2
0.1
0.1
0
X
LOAD
ATM
BRCA1
BRCA2
PALB2
E
ABUNDANCE
32768
8192
2048
512
128
64
COPY NUMBER
64
32
16
8
4
2
1
0.5
0.2
0.1
X
X
LOAD
ATM
BRCA1
BRCA2
PALB2


F
ABUNDANCE
262144
65536
16384
4096
1024
256
COPY NUMBER
32
16
8
4
2
1
0.5
0.2
0.1
0.1
0
LOAD
ATM
BRCA1
BRCA2
PALB2
G
ABUNDANCE
32768
8192
2048
512
128
64
32
COPY NUMBER
32
16
8
4
2
1
0.5
0.2
0.1
0.1
0
LOAD
ATM
BRCA1
BRCA2
PALB2
H
ABUNDANCE
131072
32768
8192
2048
512
128
COPY NUMBER
32
16
8
4
2
1
0.5
0.2
0.1
0.1
0
LOAD
ATM
BRCA1
BRCA2
PALB2


I
ABUNDANCE
16
32
64
128
512
2048
8192
COPY NUMBER
0
0.1
0.2
0.5
1
2
4
8
16
32
LOAD
ATM
BRCA1
BRCA2
PALB2
J
ABUNDANCE
32
64
128
512
2048
8192
32768
COPY NUMBER
0
0.1
0.2
0.5
1
2
4
8
16
32
X
X
LOAD
ATM
BRCA1
BRCA2
PALB2
K
ABUNDANCE
128
512
2048
8192
32768
131072
COPY NUMBER
0
0.1
0.2
0.5
1
2
4
8
16
32
LOAD
ATM
BRCA1
BRCA2
PALB2


L
M
N
ABUNDANCE
COPY NUMBER
LOAD
ATM
BRCA1
BRCA2
PALB2


O
ABUNDANCE
32768
8192
2048
512
128
64
32
COPY NUMBER
64
32
16
8
4
2
1
0.5
0.2
0.1
0.1
LOAD
ATM
BRCA1
BRCA2
PALB2
P
ABUNDANCE
8192
2048
512
128
64
32
16
COPY NUMBER
32
16
8
4
2
1
0.5
0.2
0.1
0
LOAD
ATM
BRCA1
BRCA2
PALB2

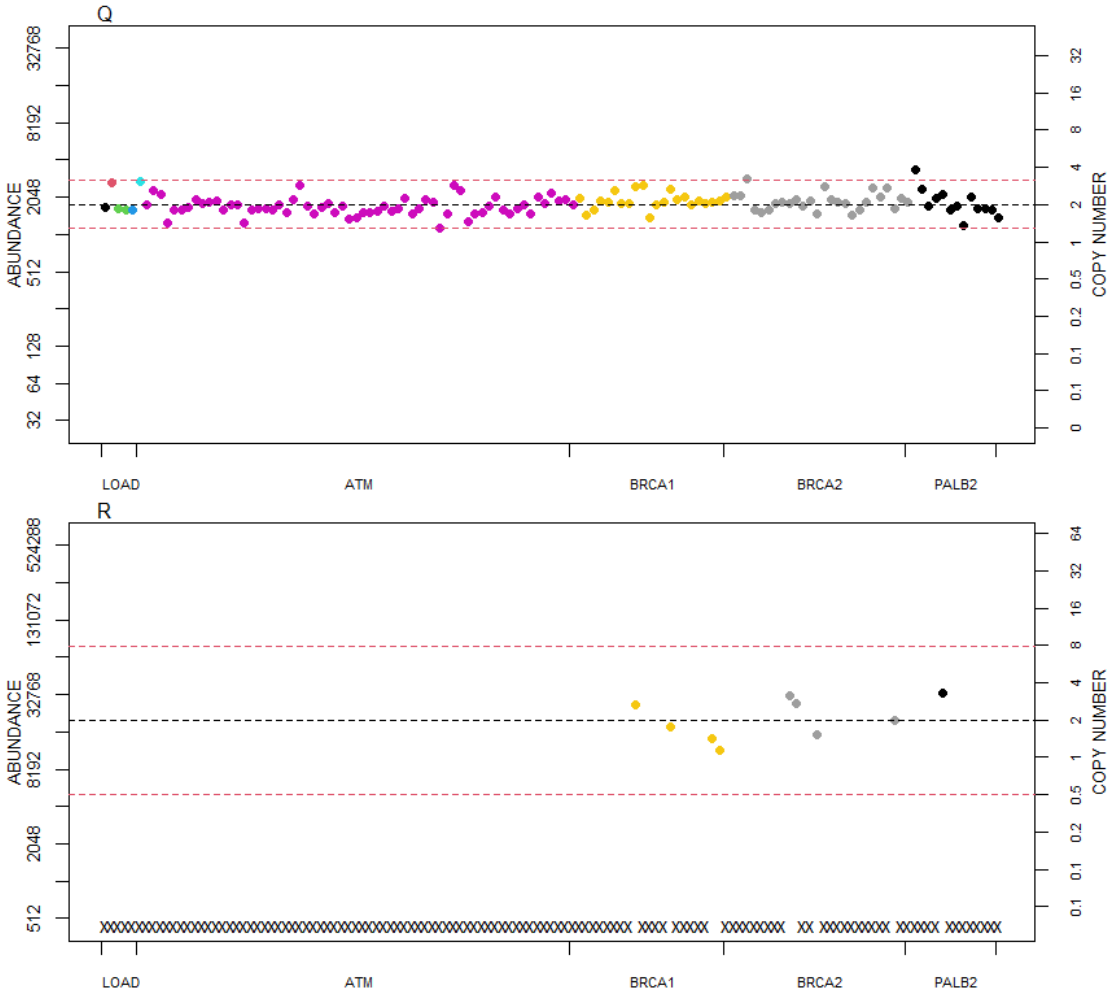

Q
ABUNDANCE
32768
8192
2048
512
128
64
32
COPY NUMBER
32
16
8
4
2
1
0.5
0.2
0.1
0
LOAD
ATM
BRCA1
BRCA2
PALB2
R
ABUNDANCE
524288
131072
32768
8192
2048
512
COPY NUMBER
64
32
16
8
4
2
1
0.5
0.2
0.1
LOAD
ATM
BRCA1
BRCA2
PALB2

Table 2. Example SNAQ™-SEQ CNV Report.

SAMPLE	VAF	ABUND	ATM	BRCA1	BRCA2	PALB2
A	4.2%	5280	1.9-2.3 (40)	1.9-2.2 (14)	0.4-0.5 (24)	1.7-2.3 (10)
B	4.6%	4280	2.9-3.5 (49)	1.3-1.6 (21)	1.3-1.6 (26)	1.8-2.2 (12)
C	30.3%	467	1.8-1.9 (54)	2.0-2.3 (22)	2.0-2.4 (25)	1.8-2.1 (13)
D	6.7%	2790	1.9-2.1 (59)	1.9-2.0 (21)	2.1-2.3 (25)	1.7-2.0 (12)
E	7.1%	2630	1.9-2.0 (59)	1.9-2.1 (22)	2.1-2.4 (25)	1.8-2.1 (12)
F	1.3%	15900	1.9-2.3 (31)	1.9-2.4 (14)	2.0-2.5 (19)	1.5-2.1 (8)
G	10.2%	1800	2.0-2.1 (60)	1.8-2.0 (21)	2.0-2.2 (24)	1.7-2.0 (11)
H	2.4%	7610	2.5-2.8 (50)	0.7-1.0 (21)	1.1-1.3 (17)	1.5-1.7 (8)
I	16.2%	995	1.8-2.0 (58)	3.1-3.5 (21)	1.1-1.3 (25)	2.6-3.0 (10)
J	12.0%	1540	2.0-2.3 (58)	1.7-2.1 (22)	0.2-0.3 (24)	1.6-1.9 (11)
K	2.7%	7300	2.5-2.8 (44)	0.7-1.0 (20)	1.3-1.4 (17)	1.3-2.0 (8)
L	5.9%	3380	2.1-2.3 (46)	0.7-0.9 (21)	1.9-2.2 (24)	1.4-1.8 (13)
M	6.3%	3000	1.9-2.2 (55)	1.6-1.9 (22)	2.2-2.6 (25)	1.4-2.0 (13)
N	8.9%	1980	1.8-1.9 (60)	1.9-2.2 (21)	2.3-2.5 (25)	1.9-2.1 (11)
O	13.7%	1300	2.3-2.7 (59)	1.0-1.5 (16)	1.8-2.4 (24)	1.4-2.3 (12)
P	19.8%	810	1.8-2.2 (57)	1.9-2.6 (22)	1.7-2.0 (24)	2.0-2.7 (12)
Q	10.0%	1790	1.8-2.0 (60)	2.1-2.3 (18)	2.0-2.2 (25)	1.8-2.2 (12)
R	1.0%	20400		1.1-2.4 (4)	1.6-3.0 (4)	NA- NA (1)

KEY

VAF indicates IS variant fraction, orange values suggest too little IS added to sample. ABUND indicates median NT gene copies, orange values suggest too little NT captured. Gene data indicate 95% confidence interval for gene copies per cell with number in () indicating number of exons used in calculation. Red indicates possible copy loss.