

CytoScan XON data analysis in Chromosome Analysis Suite (ChAS) software

A quick reference guide for CytoScan XON data analysis using ChAS software

Revision date: April 2019



Software version

For analysis of Applied Biosystems[™] CytoScan[™] XON microarray files, you must install:

- Applied Biosystems[™] ChAS software v3.3 or higher
 - Link to download ChAS v3.3 or higher
- CytoScan XON library files
- Browser/NetAffx[™] Genomic Annotations for hg19 and/or hg38 (you can only use one at a time)

Supported sample types

Default QC metrics are for postnatal blood samples and could vary for prenatal and cancer specimens.

applied biosystems

by Thermo Fisher Scientific

Suggested files to use

CytoScan XON Suite (.xnchp) sample file(s): Open (ensure that NetAffx Genomic Annotations version is correct)

XON Region Levels: On or **Off** per user preference (can use **Filters** to select gene levels 1–4)

Default Histogram/XON DGV Histogram: On (CytoScan XON_aDGV is derived from 1,855 peripheral blood samples from phenotypically normal individuals)

Genes or OMIM[®] Genes: On

AED/BED files: On if relevant to phenotype or NGS results. In addition to disease-specific AED/BED files, the following Database of Genomic Variants AED files may be useful:

• CytoScan_XON_DGV_Gains_hg19_20180521.aed CytoScan_XON_DGV_Losses_hg19_20180521.aed

Or

CytoScan_XON_DGV_Gains_hg38_20180521.aed CytoScan_XON_DGV_Losses_hg38_20180521.aed

- If desired, you can load one or more additional AED/ BED files. Right-click on each file individually and select "Include in CytoRegions". Multiple AED/BED files will be treated as one large CytoRegion, with vertical gray bars running through the data, indicating clinically significant gene regions.
- To perform truly targeted analysis, you can restrict the analysis to only those CytoRegions, and all other regions will be hidden from the view:
 - View menu > Restrict to CytoRegions
 (or button)

VCF files: On if NGS findings are relevant. (VCF files must contain only genotype or indel data.)

Files
✓ Image: State of the state
🗹 🚟 * DECIPHER_DDG2P_EMBL-EBI_2
🗹 🚟 CytoScan_XON_DGV_Gains_hg19_2
🗹 🚟 CytoScan_XON_DGV_Losses_hg19_
Derived_vcf13_19449Fprocol_vcf_report.v
XON Region Levels
🔲 🎷 DGV
🔲 🛃 Ensembl Genes
🗹 🧮 Genes
OMIM ® Genes
OMIM ® Phenotype Loci
Segmental Duplications
🔲 🕌 sno/miRNA
🔲 📘 Default Histogram
🗹 📘 🏹 XON DGV
Markers
🗹 🚥 Cytobands



Database: These aDGV XON files can be useful for comparison to segment calls in your samples:

• ChASDB_XON_aDGV_20180626.backup ChASDB_XON_aDGV_20180628_hg38.backup

Or

You may prefer to use your own database of abnormal segments in conjunction with the XON_DGV_Gains.aed and XON_DGV_Losses.aed files mentioned above.

- If you use the XON_aDGV as the database, you can hover your mouse over the histogram (Default Histogram), and it will show the number and percentage of samples in the database with that same segment.
- If you use the XON_DGV_Gains/Losses.aed files, the frequency of positive samples will appear above the segment in the AED file, but only if the frequency is greater than 1%. Thus, if you use both the database and the AED files, there may be a discrepancy in frequency.

applied blosystems

Named Setting section

Select **XON-Level 1** or user-defined **Named Setting** as shown here.

Data Types section

Gain (XON Region) and Loss (XON Region): On

Log2 Ratio: On

Summarized Log2Ratio: On

 If you have XON Merging turned On in the User Configuration, you won't see the Summarized Log2Ratio values for merged segments, because Summarized Log2Ratio is specific to each exon region.

LOH: Off, unless looking for large regions (>5 Mb)

SmoothSignal: Off, unless looking for larger regions, because there's no distinct difference for exon-level changes. If you choose to have **SmoothSignal On**, adjust the graph display to **Bar** (rather than **Line**) for easier visualization and interpretation.

Allele Difference and BAF (B Allele Frequency): Off, unless looking for large regions

Named Setting	
Karen's XON Settings	~
👪 LOH only (3Mb and 50 SNPs)	
ڏ LOH only (5Mb and 50 SNPs)	0
🍋 OncoScan Default	
🏜 Standard	
KON-Level 1	
Floor Settings	
Karen's XON Settings	
OncoScan Filter Settings	

Data Types
🗹 📈 Log2 Ratio
🔲 🖂 Weighted Log2 Ratio
🔲 📨 SmoothSignal
COH Y LOH
🔲 🖂 LOH
Allele Difference
🔲 🚧 BAF
XON Region Copy Number State
🗹 🕂 🍸 Gain (XON Region)
🗹 🕂 🍸 Loss (XON Region)
Genotype Calls
🗹 🚈 Summarized Log2Ratio

applied biosystems

by Thermo Fisher Scientific

Segments Filters **Y**

Gene levels:

- Level 1: Genes of high clinical relevance
- Level 2: ClinVar genes not in Level 1
- Level 3: Other OMIM genes
- Level 4: Other regions from RefSeq, UCSC, Ensembl, and/or Leiden Open Variation Database (LOVD)
- Start with Level 1 genes to minimize number of calls for analysis
 - In ChAS, Level 1 genes are displayed in yellow blocks
 - In ChAS, Levels 2-4 are displayed in brown blocks
- Gain:
 - Marker Count: Off
 - Size (kbp): Off
- Loss:
 - Marker Count: Off
 - Size (kbp): Off
- LOH:
 - Size (kbp): On, minimum of 5 Mb

V Genome		
Whole Region		
Hide All Segments In Ti	his Region	
K LOH		
Marker Count	0	5000
Size (kbp) 5000	0	2000
Overlap 100	100	0
✓ Level 1 🔲 Level 2 (Level 3 Level 4	
Gain (XON Region)	Cever3 Cever4	
Level 1 Level 2 Gain (XON Region) Marker Count		5000
Cevel 1 Level 2 Cevel 1 Cevel Cevel		5000
Level 1 Level 2 Gain (XON Region) Marker Count Size (kbp) Overlap 100		5000 20000 0
	0 0 0 0 0 0	5000 20000 0
Level 1 Level 2 Gain (XON Region) Marker Count Size (kbp) Overlap 100 Loss (XON Region) Marker Count		 20000 0 0
V Level 1 Level 2 Gain (XON Region) Marker Count Size (kbp) Overlap 100 Loss (XON Region) Marker Count Size (kbp) Size (kbp)		5000 20000 0 5000 20000





User Configuration 🍺

Segment Data tab:

- Use default segment data rules configuration: Off
- XON Merging: Off

DB Query tab:

Additional information on coverage and overlap functions can be found in the ChAS User Guide.

Segment Data	QC Thresholds	Color Rules	Misc	Vocabularies	DB Query	Filtered DB Query	Export
CvtoScan XON		~					
or CHP files of e	ach array type, edi appens at file load	it the configural	ion of S figuratio	moothing, Joinir	ng or Merging s before any s	to create Copy Numl Segment Filtering).	ber Segr
or CHP files of e his processing h	each array type, edi happens at file load	it the configural	tion of S figuratio	moothing, Joinir on saving (alway	ng or Merging s before any s	to create Copy Numl Segment Filtering).	ber Segr

Segment Data QC Th	resholds	Color Rul	es Misc	Voc	abularies	DB Query	Filtered DB Query	Exports
Copy Number Query P	aramete	rs						
Minimum % CN Coverage	50]	1 1 1	1 1	1.1			
Minimum % CN Overlap	50]	· · · ·	1 1	11			
	🗹 Ma	tch only sam	ne gain/los	s type	Includ	e LOH 🗹 In	clude XON Regions	
LOH Query Parameter	5							
Minimum % LOH Coverag	e 50]	· · ?	1 1	1.1			
Minimum % LOH Overlap	50]	· · · ·		- F - F			
XON Regions Query P	aramete	rs						
Minimum % Coverage	50		- · · · ·	1.1	1.1			
Minimum % Overlap	50		· · · · ·	1 1	1 1			
	🗹 Mat	tch only sam	ne gain/los	s type	Includ	e Copy Numb	per Segments	

Segments Table column selection

- Ensure that **Summarized Log2Ratio** and **DB Count Both** are in the **Chosen** section (left in the screenshot at right) for visualization in the table view.
- Set up **Table Sorting** as shown at right (or in a userdefined way).

Select Columns			×		
Choose columns to display by dragging between 'Availab Choose Chosen File Chosen Cytoband State Type Chromosome Cytoband Start Size (kbp) Marker Count Summarized Log2Ratio DB Count Both Genes Gene Count CytoRegions DB Coverage Count DB Coverage Count DB Coverage Count Chromosome Cytoband Coverage Count Chromosome Chromosome Coverage Count Chromosome Chromo	le' and 'Chosen'. Re-or Available Allele Origin BACs Call Approval Condition ID Type Condition ID Value Confidence (ReproSe Curation Time Ensembl Gene Court Ensembl Genes Filtered DB Coverage Filtered DB Coverage Filtered DB Overlap C	der entries in 'Chosen' b eq) t t count count	y dragging.		
Call	Following Marker				
Table Sorting Optionally select one or more columns for sorting data in	Filter 'Available'				
Sort by: DB Count Both	~	ASCENDING V	Clear		
Then by: Summarized Log2Ratio	~		Clear		
Then by: <pre><ro selection=""></ro></pre>	en by: Ko Selection>				
ОК	Cancel				



CytoScan XON sample analysis

Open a CytoScan XON sample file:

- 1. Look at the **QC Metrics**; thresholds are shown below.
 - MAPD ≤0.20
 - SNPQC ≥10
 - Waviness SD ≤0.08
- 2. Go to Filter Settings and turn On "Level 1 Genes" only (check box).
- 3. Go to **Whole Genome View** to see if there's anything obvious in the tracks (large **Loss/Gain**).
- 4. Go to Segments Table.
 - There are two new columns in this table:
 - Summarized Log2Ratio
 - XON Region Level
 - Note that the software doesn't give Copy Number State. It only calls Gain or Loss.
 - Note the number of calls in the upper right corner of the Segments Table:
 10 results is shown as an example
 - Sort on DB Count Both: Sort on "0"s first (large arrow up); focus on losses/ gains that are <10% of what's in the DB. In the aDGV database of 1,855 phenotypically normal samples, that would be ~18.
 - Sort on Summarized Log2Ratio: Sort on most negative values (small arrow up) for Loss; values below 0.5 might be more "real". Sort on most positive values (further down the list) for Gain. The larger the number, the more likely it is to be a true call.
- 5. Apply **Gene Lists/AED/BED** files (DECIPHER, disease-specific files, etc.), especially when there's a distinct phenotype. Right-click on the name of the file and select **Set a CytoRegion for Targeted XON analysis**.
- 6. If you've set a CytoRegion for targeted analysis (step 5), you can also restrict the view to these areas using the **Restricted Mode** button in the toolbar, or go to the **View** menu > **Restrict to CytoRegion**. This will hide all data with the exception of what's in the targeted region file.



- 7. Using Edit mode in , you can manually edit the calls to merge exon segments into one larger call, or adjust the breakpoints. You can assign a copy number if you feel confident about it, but it's not necessary. As mentioned previously, merging segments will eliminate the Summarized Log2Ratio for an exon call.
- 8. Some probes (SNPs) are used for Allele Difference or B Allele Frequency, but not for XON calls. These probes are shown as gray dots in the tracks, not as colored dots.
- 9. Smoothing and Joining are optional and should be turned Off in User Configuration.

applied biosystems

by Thermo Fisher Scientific

General comments

- For a **targeted approach**, as an orthogonal method to confirm NGS or data from other microarrays:
 - Search for a gene or region, and/or load targeted **AED/BED** file(s), and right-click on file name to create a CytoRegion for targeted XON analysis.
 - Restrict the region of analysis as described above, if desired.
 - Scan through the Loss and Gain segments in the vertical gray bars, examining the size, number of markers, number of genes, gene significance, etc., and confirm whether the calls are significant.
 - Examine any large (≥5 Mb) regions of LOH to determine relevance.
- For a **nontargeted approach**, for samples where you may be blinded to other findings, or there's no accompanying case history or phenotype:
- Ultimately, you should sort data based on **DB Count Both (ascending)** and then on **Summarized Log2Ratio (ascending)**. You can do this in two different ways:
 - Click on the Table icon at the top right corner of the Segments Table tab:
 The Select Columns table appears, in which you can drag attributes from the right (Available) column to the left (Chosen) column. You can then drag the attributes up or down to place them in the desired order in the Segments Table.
 - Alternatively, you can click on the column header and double-sort by dragging the columns to the desired location.
 - Ultimately, you should focus on those segments that have a small DB
 Count Both, but a large Summarized Log2Ratio (most positive and most negative numbers) to help gauge whether it's likely to be a true call vs. a false-positive call, or a normal copy number variant. Opening other samples (.xnchp files) at the same time can provide additional insight as to veracity.
 - Hint: For exploratory studies, use an arbitrary cutoff value of 15–18 counts for DB Count Both when using the aDGV as the database. (This would represent less than 1% of the samples in the database.)



Notes



Notes

For Research Use Only. Not for use in diagnostic procedures. © 2019 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. Online Mendelian Inheritance in Man (OMIM) is a registered trademark of the Johns Hopkins University. **COL23251 0419**