N_XCLINICAL™ ANNOTATION AND TRACK UPDATES RELEASE NOTES

FEBRUARY
2023

www.bionano.com

Effective Date: 03/03/2023

This software is designed to assist clinicians and is not intended as a primary diagnostic tool. It is each lab's responsibility to use the software in accordance with internal policies as well as in compliance with applicable regulations.

The N_xClinical™ Track Updates Release Notes document was written at BioDiscovery, LLC., a Bionano Genomics company, 715 North Douglas Street, El Segundo, CA 90245.

Document Revision: 20230216

Trademarks

Patents

Several patents are pending.

CONTENTS

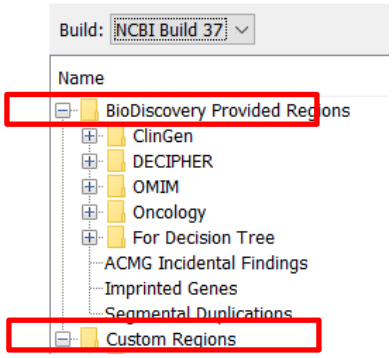
CHANGES TO TRACKS AND ANNOTATIONS

Major changes to tracks and annotation files used in N_xClinical that are provided and maintained by Bionano are outlined below. With each track update release, the N_xClinical Admin will be prompted to download, install, and deploy the new tracks to all users. Those behind a firewall and not connected to the internet will require a manual update (by their Bionano support representative) and will not receive an alert via the software.

Any change to the track file including the following will prompt an update:

- Content change at the source site – this will be incorporated into the N_xClinical track/annotation files.
- Changes to gene names/positions - Even if there is no change to the content at the source site, there may be changes introduced to the regions/annotations files due to changes in RefSeq genes. If the position of a gene changes or the name changes, the change will be introduced into the region file as well.
- Changes to track meta data – Information about the track (e.g., track description, linked URL…)
- Changes to display of track elements – Track elements may undergo a change in the color, label, or URL links.

Tracks provided and updated by Bionano are housed in a folder called “BioDiscovery Provided Regions” within the Regions tab in the UI. User-created regions (tracks) are housed in a folder called “Custom Regions.”



N_xClinical version 6.0 and newer

Only N_xClinical version 6.0 and newer will receive tracks and annotations updates.

N_xClinical version 5.2 and older

Annotations and track updates for N_xClinical 5.2 and older have been phased out and are no longer provided. Users should transition to the latest N_xClinical version as soon as possible to ensure they continue receiving the latest track/annotation updates.

Changes to Annotation Files

These are files used internally within the software; these include files for annotations, genes, and HPO terms. In this first set of official Release Notes, the content of each annotation file will be outlined. With future updates, only changes to annotation files will be documented in the Release Notes.

Genes track

Track data was obtained from UCSC Genome Browser on December 01, 2022.

Hg19: Source data version: NCBI Homo sapiens Updated Annotation Release 105.20220307 (2022-03-12)
Hg38: Source data version: NCBI Homo sapiens Updated Annotation Release 110 (2022-04-12)

DGV track

No changes. Source: 2020-02-25 release

HPO

Content was updated. Source: Release version 2022-10-05 https://hpo.jax.org/

Canonical Transcripts

Content was updated.

Changes to Regions (Tracks)

Updates below apply to both hg19 and hg38, unless indicated otherwise.

ACMG Incidental Findings Track

Source: ACMG SF v3.1. Source content has not been updated, however gene coordinates may have changed.

CIVIC Tracks

Obtained from CIViC (https://civicdb.org/) on December 01, 2022. Content in the tracks has changed as the source content is updated daily.

ClinGen Dosage Sensitivity Tracks

Downloaded on December 01, 2022, from clinicalgenome.org. Content has been updated.

ClinGen Prenatal Tracks

Source: dbVar nstd75. There were no changes to the source content and therefore no changes to the track data but the linked urls have changed slightly. The url goes to the same webpage, however how the event is displayed in the webpage's search bar has changed. .

ClinGen Postnatal Tracks

Source: dbVar nstd102. Content has been updated. Additionally, the linked url for each entry has changed slightly. The url goes to the same webpage, however how the event is displayed in the webpage's search bar has changed.

CLINGEN PRENATAL AND POSTNATAL TRACKS

These tracks were retired at the end of 2022 and will no longer be updated via the track update process. Source content for the Postnatal tracks has changed where variant calls have been aggregated into one event, making it difficult to combine prenatal and postnatal datasets. The prenatal and postnatal tracks still exist separately so you can switch to using those if you have been using the **ClinGen Prenatal and Postnatal** track set. Please contact Support with any questions.

DDG2P TRACKS

Content has been updated. Downloaded on December 01, 2022 from DDG2P database.

DECIPHER SYNDROMES TRACKS

Source: DECIPHER CNV Syndromes. No change.

IMPRINTED GENES TRACKS

Source: geneimprint.com. Content has been updated. Downloaded on December 01, 2022.

OMIM TRACKS

Data was obtained from the source on December 01, 2022, for all OMIM tracks.

New to this release, the **OMIM Syndromes** track includes additional syndromes (~45 in the current update) that are not labelled as deletion or duplication syndromes and are indicated in black color on the track. These additional entries are not represented in the **OMIM Morbid Phenotypes Deletion Syndromes** or **OMIM Morbid Phenotypes Duplication Syndromes** tracks.

REFSEQ TRACKS

Includes RefSeq Genes, RefSeq Exons, RefSeq Coding Genes, RefSeq Non-Coding Genes

Obtained from UCSC Genome Browser on December 01, 2022. Source content has been updated for hg38 but content has not been updated for hg19.

SEGMENTAL DUPLICATIONS TRACKS

Source: UCSC. No change.

TRACKS AND ANNOTATIONS CONTENT

N_xClinical has annotations from many external databases integrated within the system for use in sample review and classification. A core set of these annotation files are available as internal annotation files and region files (region files are also referred to as tracks) within the initial N_xClinical installation. Regions/tracks provided by Bionano are housed in a folder called "BioDiscovery Provided Regions" as seen in the "Regions" tab in the UI. Tracks in this folder are updated by Bionano and users (even Admins) cannot delete these tracks. Any user added tracks will be housed in the "Custom Regions" folder. Review the section on "Regions" in the "NxClinical System Administration Guide" for more information.

An update system within N_xClinical will update tracks and annotations when new content is available. The local N_xClinical Administrator is informed of any updates to these files and chooses if and when to install the updated files. Bionano aims to provide quarterly updates of these files.

ANNOTATION FILES

These are files used internally within the software; these include files for annotations, genes, and HPO terms.

Genes track

Source: NCBI RefSeq Genes (curated subset) from UCSC Genome Browser. The Genes track displays genes and transcripts obtained from RefSeq. **DGV track**

The "DGV" track in N_xClinical uses data obtained from http://dgv.tcag.ca/. The DGV track is not a replicate of the downloaded DGV data; that data is curated by Bionano to create the N_xClinical DGV track. N_xClinical version 6.0 and later does not include the **DGV Gold Standard** tracks. The DGV track in N_xClinical is filtered to provide a higher quality set of variants than what is provided in the source DGV data download. It is curated to include only high-quality results according to the following filtering criteria:

- exclude CNVs from any study that includes BAC results
- exclude all events less than 50bp (to filter out seq var changes from sequencing data e.g., gnomAD-SV)

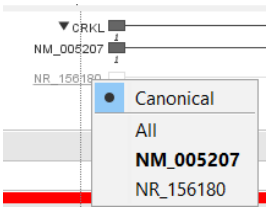
The DGV Gold Standard track can accessed from http://dgv.tcag.ca/ and uploaded into N_xClinical as a custom track if desired.

HPO (Human Phenotype Ontology)

HPO terms and gene to phenotype mapping are obtained from https://hpo.jax.org/.

Canonical Transcripts

The canonical transcripts file is used to select the default transcript for display in the Genes track. E.g.



Genes from the N_xClinical genes track are matched to locate the canonical transcript from the data sources listed below. If the gene is not found, the next data source in the list is searched, and so on until all genes in the Genes track have been queried. There are a small number of genes that could not be located in the sources below. In such cases, N_xClinical will display the longest transcript in the Genes track.

Databases queried, in order, to obtain canonical transcripts:

Hg19:

1. RefSeq Select
2. VEP version 99

Hg38:

1. MANE Select v1.0
2. RefSeq Select
3. VEP version 99

TRACKS/REGION FILES

N_xClinical Regions (tracks) are composed of sets of tracks provided and maintained by Bionano as well as user-loaded tracks. Only the regions provided by Bionano are updated via the Bionano tracks and annotations update process.

The track description contains information of the source content, summary of track data, and the downloaded date. Hover over the track name in the Table and Track Preferences to view these details.

ACMG INCIDENTAL FINDINGS TRACK

The track contains the updated gene list of 78 genes (ACMG SF v3.1) published in June 17, 2022. (https://www.gimjournal.org/article/S1098-3600(22)00723-7/fulltext). This is the minimal list of genes recommended by the American College of Medical Genetics and Genomics (ACMG) to be reported as incidental or secondary findings.

CIVIC TRACKS

Obtained from CIViC (https://civicdb.org/). Contains cancer related variants from the CIViC knowledge base. Only variants records that have been accepted are included here.

Note that CIViC only provides hg19 coordinates. For the hg38 track, the coordinates were converted using the NCBI remap tool.

CIViC Variants track

This track contains variants from CIViC. The contents (variants in cancer genes with the event label in the track displaying the gene name and variant) are the same. The contents of this track comes from the "Variant Summaries" file downloaded from the CIViC website. The "Variant Summaries" file only contains variants with accepted variant records.

CIViC Genes track

The **CIViC Genes** track now contains only genes implicated in cancer as found in the "Gene Summaries" file downloaded from the CIViC website. This track contains all genes from the CIViC website regardless of the status of variant records. Even if a gene does not have any accepted variant records, the gene is present in this file.

CLINGEN DOSAGE SENSITIVITY TRACKS

Data is obtained from ClinicalGenome.org. Dosage Sensitivity Map genes and regions for which evidence was collected either supporting or refuting haploinsufficiency or triplosensitivity. Evidence in the ClinGen Dosage Sensitivity curation process is evaluated on a continual basis by the ClinGen Structural Variation Working Group as described in Riggs et al. 2012 (http://www.ncbi.nlm.nih.gov/pubmed/22097934). The results are compiled into the ClinGen Genome Dosage Map Resource.

Main tracks:

- ClinGen Dosage Sensitive Map All
- ClinGen Dosage Sensitive Map Benign
- ClinGen Dosage Sensitive Map Likely Pathogenic
- ClinGen Dosage Sensitive Map Pathogenic

Decision Tree tracks with folder hierarchy:

- Gains
    - Genes
        - ClinGen Dosage Sensitive Map Triplosensitivity Benign Genes
        - ClinGen Dosage Sensitive Map Triplosensitivity Likely Pathogenic Genes
        - ClinGen Dosage Sensitive Map Triplosensitivity Pathogenic Genes
    - Regions
        - ClinGen Dosage Sensitive Map Triplosensitivity Benign Regions
        - ClinGen Dosage Sensitive Map Triplosensitivity Likely Pathogenic Regions
        - ClinGen Dosage Sensitive Map Triplosensitivity Pathogenic Regions
- Losses
    - Genes
        - ClinGen Dosage Sensitive Map Haploinsufficiency Benign Genes

        - ClinGen Dosage Sensitive Map Haploinsufficiency Canonical Transcript
        - ClinGen Dosage Sensitive Map Haploinsufficiency Gene Components
        - ClinGen Dosage Sensitive Map Haploinsufficiency Likely Pathogenic Genes
        - ClinGen Dosage Sensitive Map Haploinsufficiency Pathogenic Genes
    - Regions
        - ClinGen Dosage Sensitive Map Haploinsufficiency Benign Regions
        - ClinGen Dosage Sensitive Map Haploinsufficiency Likely Pathogenic Regions
        - ClinGen Dosage Sensitive Map Haploinsufficiency Pathogenic Regions

ClinGen Dosage Curation scoring:

- 3: Sufficient evidence suggesting dosage sensitivity is associated with clinical phenotype
- 2: Emerging evidence suggesting dosage sensitivity is associated with clinical phenotype
- 1: Little evidence suggesting dosage sensitivity is associated with clinical phenotype
- 0: No evidence to suggest that dosage sensitivity is associated with clinical phenotype
- 40: Dosage sensitivity unlikely
- 30: Gene associated with autosomal recessive phenotype

Clinical classification is extrapolated from the scores. Segmented tracks are created for benign, likely pathogenic, and pathogenic based on the scores above. 3=pathogenic; 1,2=likely pathogenic; 40=benign.

0 and 30 are ignored and not included in our Decision Tree tracks.

The tracks below are used by the ACMG Scoreboard feature:

- ClinGen Dosage Sensitive Map Haploinsufficiency Canonical Transcript
- ClinGen Dosage Sensitive Map Haploinsufficiency Gene Components

Coloring for track events:

- Loss events -> red
- Gain events -> blue

CLINGEN PRENATAL TRACKS

Source: dbVar nstd75. Contains CNVs in the prenatal array dataset from Wapner et al. 2012 (http://www.ncbi.nlm.nih.gov/pubmed/23215555), also known as nstd 75 (ClinGen Prenatal).

Main tracks:

- ClinGen Prenatal All
- ClinGen Prenatal Benign
- ClinGen Prenatal Likely Benign
- ClinGen Prenatal Likely Pathogenic
- ClinGen Prenatal Pathogenic
- ClinGen Prenatal Uncertain Significance

Decision Tree tracks:

- ClinGen Prenatal Gains Benign
- ClinGen Prenatal Gains Likely Benign
- ClinGen Prenatal Gains Likely Pathogenic
- ClinGen Prenatal Gains Pathogenic
- ClinGen Prenatal Gains Uncertain Significance
- ClinGen Prenatal Losses Benign
- ClinGen Prenatal Losses Likely Benign
- ClinGen Prenatal Losses Likely Pathogenic
- ClinGen Prenatal Losses Pathogenic
- ClinGen Prenatal Losses Uncertain Significance

Coloring for track events:

- Loss events -> red
- Gain events -> blue

CLINGEN POSTNATAL TRACKS

Source: dbVar nstd102. Track contains CNVs designated as benign, likely benign, pathogenic, likely pathogenic, conflicting, and regions of uncertain significance in clinical structural variants submitted to ClinVar. It replaces and supplements clinical variants from several studies that were originally submitted to dbVar (e.g., nstd37, nstd101).

How Variant Calls are represented in nstd102 and our tracks

- Variant calls in the study are aggregated into one entry per call type. If all classifications match, then there will be a single top-level aggregate classification. If they do not match, the classification will be "Conflicting." E.g., if there were five cases with the following:
    - Gain-benign
    - Gain-benign
    - Gain-benign
    - Loss-Pathogenic
    - Loss-Benign

  The study consolidates these into the following two entries:
    - Gain-benign
    - Loss-Conflicting interpretations of pathogenicity
- In addition to segmented tracks based on clinical interpretation (**Pathogenic**, **Likely Pathogenic**, **Uncertain Significance**, **Likely Benign**, and **Benign)**, there is an additional segment called **Conflicting**. The **Conflicting** tracks contain events in nstd102 with interpretation values that do

not fall into one of the segments listed above. Values such as the following would be segmented into the **Conflicting** track:
    - Conflicting data from submitters
    - Conflicting interpretations of pathogenicity
- If there is more than one clinical interpretation in an entry classification, the entry will be segmented into each individual clinical interpretation track.
    - e.g. “Benign/Likely benign:See cases” will be segmented into the Benign track and Likely benign track.
    - e.g. “Uncertain significance, Pathogenic/Likely pathogenic:See cases” will be segmented into the Uncertain significance track, Pathogenic track, and Likely pathogenic track.
- Call types **copy number gain** and **duplication** in the study = **gain** in our tracks
- Call types **copy number loss** and **deletion** in the study = **loss** in our tracks
- The label displays clinical significance in addition to the subject phenotype. E.g., Pathogenic: Developmental Delay
- The “Subject Phenotype” field contains values such as “See Cases” or “not provided” for most of the events, therefore many entries may have labels such as “Pathogenic: See Cases”.

Coloring for track events:

- Loss events -> red
- Gain events -> blue

DDG2P TRACKS

Source: DDG2P database. A curated list of genes reported to be associated with developmental disorders in the Decipher Developmental Disorders Gene2Phenotype database (DDG2P).

Includes all DDG2P related tracks in N_xClinical. Note: DECIPHER has moved to hg38 (Dec. 2020). Both hg19 and hg38 track coordinates are obtained via DECIPHER.

The tables below show how the events are segmented into respective “For Decision Tree” tracks based on “DDD category,” “confidence” and “allelic requirement” labels in the source data.

NxC Track name contains	**Confidence category**	**D2P Allelic Requirement**
Biallelic Confirmed	Definitive	biallelic_autosomal biallelic_PAR
Monoallelic Confirmed	Definitive	monoallelic_autosomal monoallelic_X_hem monoallelic_Y_hem monoallelic_X_het

		monoallelic_PAR
Biallelic Unconfirmed	Strong Limited Moderate Both RD and IF	biallelic_autosomal biallelic_PAR
Monoallelic Unconfirmed	Strong Limited Moderate Both RD and IF	monoallelic_autosomal monoallelic_X_hem monoallelic_Y_hem monoallelic_X_het monoallelic_PAR
Recessive	Definitive Strong Limited Moderate Both RD and IF	biallelic_autosomal biallelic_PAR
Dominant	Definitive Strong Limited Moderate Both RD and IF	monoallelic_autosomal monoallelic_X_hem monoallelic_Y_hem monoallelic_X_het monoallelic_PAR

Coloring for DDG2P track events:

- Dominant/monoallelic events -> green
- Recessive/biallelic events -> orange

DECIPHER SYNDROMES TRACKS

Source: DECIPHER CNV Syndromes. DECIPHER CNV Syndromes is a list of expert-curated microdeletion and microduplication syndromes involved in developmental disorders.

Note that on the source site, a few events do not have coordinates in build 37 because DECIPHER moved to display of coordinates in hg38 (end of 2020); coordinates were lifted over to hg38. For some events, manual annotation was required either because events could not be lifted over automatically or there was a problem with the lifted over coordinates. DECIPHER does not provide hg19 coordinates for these events since the events were not lifted over automatically. For these handful of events, we have used the build 37 coordinates provided by DECIPHER prior to their change to display of events in hg38.

Coloring for track events is as follows:

- deletion syndrome -> red
- duplication syndrome -> blue

IMPRINTED GENES TRACKS

Source: geneimprint.com. A curated list of gene names, status (predicted, imprinted, etc.) and the expressed allele (maternal, paternal, isoform dependent, unknown, etc.) in the imprinted gene database (Gene Imprint). This track excludes all entries with status=Not Imprinted.

The “Imprinted Genes” track includes all events from the Gene Imprint db except those labeled as “Not_Imprinted”. The default link from this track is to OMIM. If the gene is not in OMIM at the time of track generation, then the entry links out to NCBI.

Coloring for track events is as follows:

- paternally expressed -> blue
- maternally expressed -> pink
- all others -> gray

OMIM TRACKS

Source: omim.org. All genes represented in OMIM (Online Mendelian Inheritance in Man), an online catalog of human genes and genetic disorders. Includes all OMIM related tracks in N_xClinical. The OMIM API is used to obtain data and only hg38 genomic coordinates are provided by OMIM. Locations for build 37 tracks are obtained from RefSeq Genes, curated subset by mapping gene names. For genes in OMIM data that are not in RefSeq curated, the positions are obtained from UCSC Genes. For build 37 OMIM Syndromes, the provided cytoband location is used and the corresponding hg19 coordinates for the cytoband are displayed. OMIM Syndromes in build 38 uses the OMIM provided genomic coordinates therefore variations may be seen between hg19 and hg38 where hg38 may have more precise coordinates than hg19.

OMIM Genes - All genes represented in OMIM (Online Mendelian Inheritance in Man), an online catalog of human genes and genetic disorders.

OMIM Morbid Phenotypes - Cytogenomic locations of the genes associated with disorders (OMIM Morbid Map) as well as phenotypes and mode of inheritance (where available) from OMIM.

Event coloring for OMIM tracks (excluding Syndromes tracks):

- Dominant events -> green
- Recessive events -> orange
- Morbid genes -> black
- All other events -> gray

OMIM Syndromes - Cytogenomic locations of deletion and duplication syndromes represented in OMIM.

OMIM Syndromes includes syndromes not labeled as deletion or duplication syndromes. These additional entries are not represented in the **OMIM Morbid Phenotypes Deletion Syndromes** or **OMIM Morbid Phenotypes Duplication Syndromes** tracks.

For build 37 OMIM Syndromes, the provided cytoband location is used and the corresponding hg19 coordinates for the cytoband are displayed. OMIM Syndromes in build 38 uses the OMIM provided genomic coordinates therefore variations may be seen between hg19 and hg38 where hg38 may have more precise coordinates than hg19.

OMIM Syndromes coloring:

- Duplication syndromes -> blue
- Deletion syndromes -> red
- Syndromes not labeled as deletion or duplication -> black

REFSEQ TRACKS

Includes RefSeq Genes, RefSeq Exons, RefSeq Coding Genes, RefSeq Non-Coding GenesSource: NCBI RefSeq Genes (curated subset) from the UCSC Genome Browser. This is the same source as that for the Genes track.

RefSeq Genes - Locations of official RefSeq genes (NCBI RefSeq genes, curated subset)

RefSeq Exons - Locations of exons of the RefSeq genes (NCBI RefSeq genes, curated subset).

RefSeq Coding Genes and **RefSeq Non-Coding Genes** - Genes that are annotated as coding genes or non-coding genes, respectively, by the GENCODE project in V42 (Ensembl 108) release. The gene position is the same as that in the Genes track (NCBI RefSeq curated subset obtained from UCSC Annotation Downloads). GENCODE is used only to label genes as coding/non-coding. The GENCODE Genes show high-quality manual annotations merged with evidence-based automated annotations across the entire human genome generated by the GENCODE project. The annotation was carried out on genome assembly GRCh38.

SEGMENTAL DUPLICATIONS TRACKS

Source: UCSC.

A summary of large genomic duplications (>1kb, >90% similar) in the genome. Detailed track description at https://genome.ucsc.edu/cgi-bin/hgTables. For a region to be included in the track, at least 1 Kb of the total sequence (containing at least 500 bp of non-RepeatMasked sequence) had to align and a sequence identity of at least 90% was required.

REGION FILE UPDATE PROCEDURE

Bionano routinely provides updated files to N_xClinical Administrators. Bionano checks the external database sites for any updates. If there are updates to the database, Bionano obtains the files and processes them for use within N_xClinical and tests and validates the files before uploading onto the update server. The N_xClinical system checks this server on a regular basis for any updates; if updates are available, the N_xClinical Administrator is alerted when theye log into N_xClinical.

If there are no updates to the external databases, Bionano does not post new files to the update server and hence no alert goes out to the N_xClinical Administrator. If a region file in N_xClinical has not been updated in a while, it's because the external database has not been updated as Bionano checks quarterly for any updates.

Region files have two main components to which changes may occur:

- Actual content (genes, locations, classifications, etc.) in the source database
- Meta properties which are the meta data for the region file (Bionano version for the region file, description of the content in the file, download date, etc.)

Sometimes there may be changes to the meta properties of the region files resulting in an available update even though no clinical/genomic content has changed within the source files.

UPDATE PROCESS FOR REGION FILES

1. Bionano checks the external database to see if there are any updates
2. If there are updates
    a. Bionano obtains the files and processes them for use within N_xClinical
    b. A version number is assigned to each region file for tracking and auditing
    c. The region properties file is updated to indicate
        i. the date the data was downloaded by Bionano from the external database
        ii. the date the external database itself was last updated
    d. Bionano validates the new region files:
        i. Reviews to make sure the source file retrieval and resulting file creations worked correctly and performs some spot checking of a few genes/regions in the region/annotation files
        ii. Tests the new region/annotation files in N_xClinical
        iii. Tests the update system with the new files
    e. Customers are notified about upcoming annotation and track updates via email. Sites connected to the internet will also get a notification within N_xClinical.
3. If the external database has not been updated since Bionano last obtained the files, region files will have no content changes but may have an update available depending on changes to the meta properties of the region file.
    a. If the meta properties (description/version) is not updated, no update for that region file will be available. Meta properties include source database version and download

date or date checked for new content; but these are not updated each time to prevent triggering unnecessary region file updates

b. If meta properties is updated, the region file will have an update available
c. Roughly once a year for each region file the meta properties will be updated to reflect that the source database was checked and there was no updated content available.

UPDATE PROCESS FOR INTERNAL ANNOTATION FILES

Internal annotation files are integral to the system and are provided with each N_xClinical installation. These include RefSeq genes, genome assembly, DGV, and HPO terms. The same procedure as that for region files is followed for updating these files.