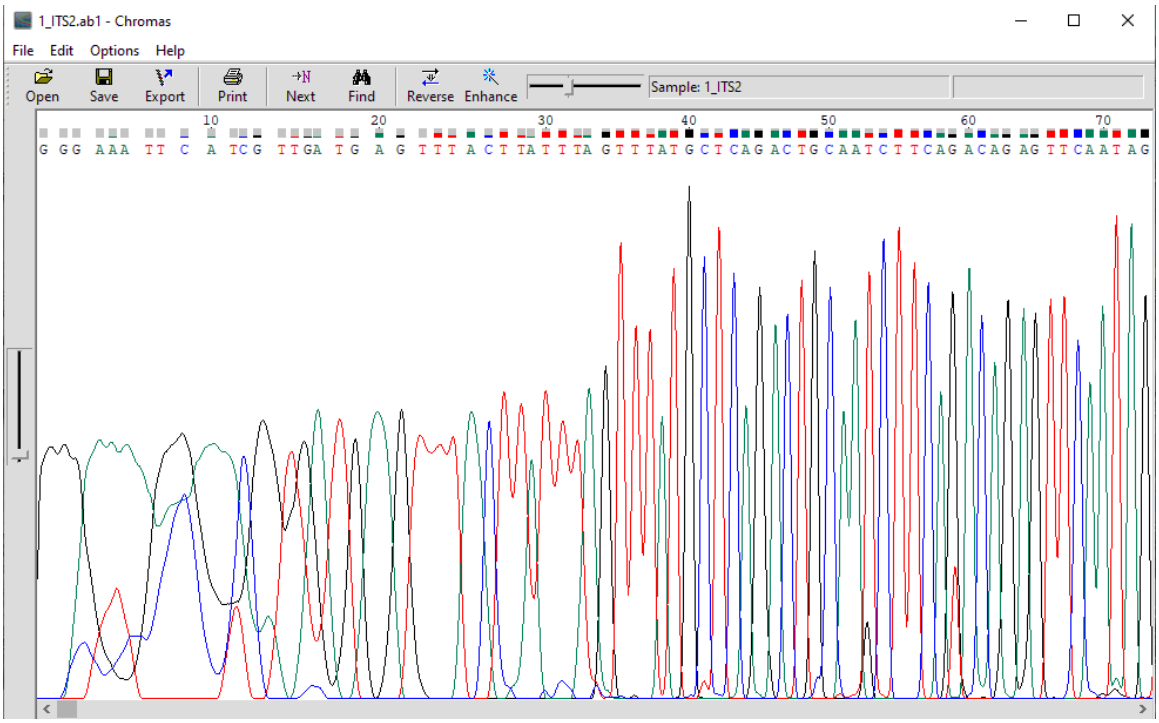
Sequence Processing:

DNA sequence processing refers to the computational and experimental steps taken to analyze and interpret the order of nucleotide bases (A, T, C, G) within a DNA molecule after it has been sequenced. This involves assessing the quality of the raw data, cleaning it up, and then applying various bioinformatics tools to extract meaningful biological information. Several software can perform this issue like R, Chromas, BioEdit, SnapGene, FinchTV, ….etc. we'll focus on chromas as it is easy to use, free, and active software.

The goal of the lecture:

1. Loading .ab1 sequence file.
2. Pre-process it (cleaning, and simple editing).
3. Save it as FASTA or FASTq file to be used in downstream processes.

Chromas software



Chromas is a popular and long-standing software used by researchers to **view, edit, and analyze DNA sequencing chromatograms**, which are the graphical representations of DNA sequences produced by Sanger sequencing. It's developed by Technelysium Pty Ltd.

How to download Chromas:

For the Free Chromas:

- Go to the official Technelysium website: https://technelysium.com.au/wp/contact/
- Navigate to the "Chromas" section.
- Look for a download link for the latest free version of Chromas (e.g., "Download Chromas").
- Click the link to download the installer file (usually a .exe file for Windows).

For Chromas Lite (Free):

- You can often find Chromas Lite on software download websites like Software Informer: https://chromas-lite.software.informer.com/2.1/.

Install Chromas (Windows):

- Once the installer file (.exe) is downloaded, locate it on your computer (usually in your "Downloads" folder).
- **Double-click** the installer file to begin the installation process.
- You may see a security warning; click "Run" or "Yes" to proceed.
- Follow the on-screen instructions in the installation wizard:
    - Read and accept the license agreement.
    - Choose the installation location (the default is usually fine).

- Decide if you want to create a desktop shortcut.
- Click "Install" to begin the installation.

- Once the installation is complete, click "Finish" to close the wizard.
- You should now be able to find and run Chromas from your Start Menu or desktop shortcut (if you created one).

Install Chromas (macOS):

- The download for macOS will likely be a .dmg (Disk Image) file.
- **Double-click** the .dmg file to mount it.
- A Finder window will open, showing the Chromas application icon.
- Drag the Chromas application icon to your "Applications" folder.
- Once copied, you can eject the mounted disk image.
- You can now run Chromas from your "Applications" folder.

Key Features of Chromas:

- **Chromatogram Viewing:** It can open and display chromatogram files in various formats, primarily .ab1 (from Applied Biosystems sequencers) and .scf. It can also open .ztr files.
- **Sequence Editing:** Allows users to edit the base calls in the sequence based on the chromatogram peaks. This includes:
  - Changing incorrect base calls.
  - Deleting bases.
  - Inserting bases.
  - Trimming low-quality regions at the start and end of the sequence based on peak quality.
- **Quality Assessment:** Displays quality scores associated with each base call (if available in the file) as colored bars above the sequence. This helps in identifying unreliable regions.

- **Sequence Manipulation:**
  - Reverse complementing the sequence and chromatogram.
  - Searching for subsequences (exact matching or optimal alignment).
  - Displaying translations of the DNA sequence in three reading frames.
- **Exporting Sequences:** Supports exporting sequences in various text-based formats like:
  - Plain text
  - FASTA
  - FASTQ
  - EMBL
  - GenBank
  - GCG
  - Can also save back to .scf or .ab1 formats.
- **Copying Data:** Allows copying the sequence to the clipboard in plain text, FASTA, or FASTQ format. You can also copy an image of a chromatogram section.
- **Batch Processing:** Offers batch processing capabilities for tasks like format conversion, sequence export with vector removal, batch printing, and batch export of raw data.
- **Vector Sequence Detection (in some versions):** Can help identify and remove vector sequences.
- **Integration with External Tools:** Can often be used in conjunction with other bioinformatics tools like BLAST for sequence identification.

- **PeakTrace Integration (paid service):** Offers an option to enhance .ab1 files using Nucleics' PeakTrace RP component to improve peak readability and extend the number of high-quality bases.

Versions of Chromas:

- **Chromas (Free):** The basic version, suitable for simple sequencing projects that don't require assembly of multiple sequences.
- **Chromas Lite (Free):** Seems to be an older or very basic version with similar functionalities to the free Chromas.
- **ChromasPro (Paid):** A more advanced version that includes features for assembling overlapping sequence reads into contigs, a graphical contig editor, restriction enzyme mapping, ORF finding, and BLAST submission.

Overall:

Chromas is a straightforward and user-friendly tool, particularly for researchers working with Sanger sequencing data who need to visualize and perform basic editing and analysis of their chromatograms. While its interface is sometimes described as a bit outdated, its core functionalities are well-implemented and responsive.

General overview of what you'd typically find in each menu of Chromas:

1. File Menu:

- **Open:** Allows you to open chromatogram files (typically .ab1, .scf, .ztr).
- **Save As:** Enables you to save the currently open file, potentially in a different format (.scf or .ab1).
- **Print:** Lets you print the displayed chromatogram and sequence, often with options to adjust the zoom level or fit it to a page.

- **Export Sequence:** Allows you to export the DNA sequence in various text-based formats:
    - **Plain Text:** Just the sequence.
    - **FASTA:** A standard format with a header line.
    - **FASTQ:** A format that includes quality scores.
    - **EMBL:** A European Molecular Biology Laboratory format.
    - **GenBank:** A National Center for Biotechnology Information format.
    - **GCG:** A format used by the GCG Wisconsin Package.
    - **Formatted with Base Numbering:** Includes numbering for easy reference.
- **Batch Processing:** Opens a dialog for performing actions on multiple files at once, such as:
    - **Format Conversion:** Converting multiple files to a different format.
    - **Sequence Export with Vector Removal:** Exporting sequences while attempting to remove vector contamination (if vector information is present or configured).
    - **Batch Printing:** Printing multiple chromatograms.
    - **Batch Export of Raw Data:** Exporting the raw data from the chromatogram files.
- **Exit:** Closes the Chromas application.

2. Edit Menu:

- **Copy Sequence:** Copies the displayed DNA sequence to the clipboard in various formats (plain text, FASTA, FASTQ).

- **Copy Chromatogram:** Copies an image of the currently displayed section of the chromatogram to the clipboard, allowing you to paste it into documents or presentations.
- **Reverse Complement:** Reverses the DNA sequence and also displays the corresponding reverse complement of the chromatogram peaks.
- **Find:** Opens a dialog to search for a specific subsequence within the DNA sequence using exact matching or optimal alignment.
- **Select All:** Selects the entire DNA sequence.
- **Trim Left:** Removes the sequence and chromatogram data to the left of the current cursor position.
- **Trim Right:** Removes the sequence and chromatogram data to the right of the current cursor position.
- **Automatic Trim (if quality data is available):** Automatically removes low-quality bases from the ends of the sequence based on the quality scores associated with each base call.

3. View Menu:

- **Zoom In:** Increases the magnification of the chromatogram display.
- **Zoom Out:** Decreases the magnification of the chromatogram display.
- **Zoom to Fit:** Adjusts the zoom level so that the entire chromatogram fits within the window.
- **Show Quality:** Toggles the display of quality scores (if present in the file) as colored bars above the sequence.
- **Show Translation:** Displays the possible protein translations of the DNA sequence in three forward reading frames below the nucleotide sequence.
- **Scroll to Base:** Allows you to jump to a specific base number in the sequence.

4. Tools Menu:

- **Options:** Opens a dialog where you can configure various settings for Chromas, such as:
  - Default save formats.
  - Printing options.
  - Display preferences.
  - **PeakTrace RP Options (if available):** If you have access to the paid PeakTrace service, this is where you might configure its use.
- **BLAST Search:** Allows you to select the current DNA sequence and initiate a BLAST (Basic Local Alignment Search Tool) search against the NCBI (National Center for Biotechnology Information) database to identify similar sequences

5. Help Menu:

- **Help Topics**: Opens the Chromas help documentation, providing information on how to use the various features of the software.
- **About Chromas**: Displays information about the Chromas version, copyright, and developer.

Steps to save a processed .ab1 file as .fasta or .fastq in the free version of Chromas:

1. Load and Process Your .ab1 File:

- Follow the steps outlined in the previous response to **load** your .ab1 file into Chromas.
- Perform any necessary **editing** of the sequence, such as correcting base calls or trimming low-quality regions. Ensure the sequence displayed in Chromas is the final, processed version you want to save.

2. Export the Sequence:

- Click on the **"File"** menu in the top left corner of the Chromas window.
- Select **"Export Sequence..."** from the dropdown menu. This will open a submenu with different export formats.

3. Choose the Desired Format:

- **To save as .fasta:**
  - In the "Export Sequence..." submenu, click on **"FASTA"**.
  - A "Save As" dialog box will appear.
  - Choose the directory where you want to save the file.
  - Enter a **filename** for your sequence file (it's common to use the original filename with a .fasta or .fa extension).
  - Click the **"Save"** button.
- **To save as .fastq:**
  - In the "Export Sequence..." submenu, click on **"FASTQ"**.
  - A "Save As" dialog box will appear.
  - Choose the directory where you want to save the file.
  - Enter a **filename** for your sequence file (it's common to use the original filename with a .fastq extension).
  - Click the **"Save"** button.

What to Expect in the Exported Files:

- **.fasta (.fa):** This format will contain:
  - A header line starting with a ">" symbol, usually followed by the original filename or a sequence identifier.
  - The processed DNA sequence on subsequent lines.

>your_filename
ACGTACGTACGT...

- **.fastq:** This format will contain four lines for each sequence:

- A header line starting with a "@" symbol, usually followed by the original filename or a sequence identifier and potentially some additional information.
- The processed DNA sequence.
- A line starting with a "+" symbol, which can sometimes be followed by the same identifier as the first line.
- A line of quality scores, where each character represents the quality of the corresponding base in the sequence. The encoding of these quality scores can vary (e.g., Sanger, Solexa, Illumina). Chromas typically exports Sanger-style quality scores if they are present in the .ab1 file.

@your_filename
ACGTACGTACGT...
+
IIIIIIIII...

Important Considerations:

- **Quality Scores in FASTA:** The FASTA format **does not** store quality scores. If you save as FASTA, you will lose the quality information associated with your sequence.
- **Quality Scores in FASTQ:** The FASTQ format **does** store quality scores, making it the preferred format if you need to retain this information for downstream analysis (e.g., alignment, variant calling).
- **Filename:** Choose descriptive filenames for your exported files so you can easily identify them later.
- **Directory:** Select a logical directory on your computer to save the exported files.

By following these steps, you can easily save your processed .ab1 files from Chromas into the widely used .fasta or .fastq formats for further bioinformatics analysis. Remember to choose the format that best suits your downstream applications.