

# GenEditScan-GUI User Guide

Version 1.0.2

December XX, 2024

National Agriculture and Food Research Organization (NARO)

国立研究開発法人農業・食品産業技術総合研究機構

GenEditScan-GUI is a tool that counts fragments of vector sequences contained in NGS data and performs a test of statistical significance. It has a function to illustrate the count number of the sequence fragment and the test results. Speeding up of calculation is achieved by bit operation and parallel processing. GenEditScan-GUI is developed by JavaFX which is one of the standard development environment of Java, and works on Windows and Mac.

## Contents

1. Introduction .....	1
2. Getting started.....	1
3. Function .....	1
4. Input and output files.....	2
5. Statistics file .....	3
6. Outside file .....	4
7. Mer frequency file .....	5
8. Count mer.....	6
9. Draw graph .....	8
9.1. グラフ表示機能 .....	8
9.2. 周辺配列の表示機能 .....	13
10. Menu.....	15
10.1. File .....	15
10.2. Help.....	16
11. License.....	17
12. Open source licenses.....	17
13. Release notes.....	19

## 1. Introduction

GenEditScan-GUI は、NGS データに含まれるベクター配列の断片をカウントし、統計的有意性の検定を行うツールである。配列断片のカウント数や検定結果を図示する機能を備えており、ビット演算と並列処理により計算の高速化が図られている。GenEditScan-GUI は Java の標準的な開発環境の一つである JavaFX により開発されており、Windows および Mac で動作する。

## 2. Getting started

### 【Windows】

「GenEditScan\_win.vbs」またはこの vbs ファイルのショートカットを作成しダブルクリックする（vbs ファイルのコピーでは動作しない）。

### 【Mac】

ターミナルから「GenEditScan\_mac.sh」を起動するか、「GenEditScan\_mac.command」をダブルクリックする。

## 3. Function

GenEditScan-GUI の機能の概要を Table 1 に示した。「Count mer」で Mutant と Wild type の両方のファイルを指定した場合、mer のカウント後に G 検定を行い、カウント数と共に検定結果をグラフ表示する。Mutant または Wild type の一方を与えた場合、G 検定は行わず、カウント数のみをグラフ表示する。

Table 1 GenEditScan-GUI の機能の概要

Tab menu	機能の概要	
Count mer	Mer（配列断片）をカウントして G 検定を行う。ベクターファイル (fasta)、Mutant ファイル (fastq)、Wild type ファイル (fastq)、K-mer（断片長）、計算スレッド数を指定する。	
Draw graph	グラフ表示機能	Mer のカウント数や G 統計量をグラフ表示する。P 値や FDR、Bonferroni への表示の切り替えが可能である。軸の範囲やタイトル、グラフの色が指定でき、PDF および PNG ファイルへの保存機能を持つ。
	周辺配列の表示機能	FDR の指定閾値を超えた配列をマウス操作により選択し、周辺配列のパターンの頻度を Mutant と Wild type 毎にカウントする。

#### 4. Input and output files

GenEditScan-GUI のタブメニュー毎の入出力ファイルを Table 2 に、ファイルの内容を Table 3 に示した。Count mer では、ユーザーが入力ファイルを用意する必要がある。表中の 3 種類全ての入力ファイルを「Count mer」に与えることで、mer のカウント後に G 検定が引き続き実行される。作成されるファイル名には GUI で指定した prefix が使われる。

**Table 2 GenEditScan-GUI の入出力ファイル**

Tab menu	入力ファイル	出力ファイル
Count mer	<ul style="list-style-type: none"><li>・ ベクターファイル (fasta)</li><li>・ Mutant ファイル (fastq)</li><li>・ Wild type ファイル (fastq)</li></ul>	<ul style="list-style-type: none"><li>・ out_prefix.statistics.txt</li><li>・ out_prefix.outside.txt</li><li>・ out_prefix.mutant.merFreq.txt</li><li>・ out_prefix.wildtype.merFreq.txt</li></ul>
Draw graph	<ul style="list-style-type: none"><li>・ out_prefix.statistics.txt</li><li>・ out_prefix.outside.txt</li></ul>	<ul style="list-style-type: none"><li>・ PDF ファイル</li><li>・ PNG ファイル</li></ul>

**Table 3 入出力ファイルの内容**

ファイル	内容
ベクターファイル	ベクター配列が格納されている fasta ファイル。ユーザーが用意する。
Mutant ファイル	Mutant の fastq (または fastq.gz) 形式のファイル。ユーザーが用意する。
Wild type ファイル	Wild type の fastq (または fastq.gz) 形式のファイル。ユーザーが用意する。
out_prefix.statistics.txt	G 統計量の結果が mer のカウント数と共に格納される。
out_prefix.outside.txt	検出された k-mer 配列周辺の解析結果ファイル
out_prefix.mutant.merFreq.txt	塩基配列と出現数 (Mutant)
out_prefix.wildtype.merFreq.txt	塩基配列と出現数 (Wild type)
PDF ファイル	Draw graph タブの「Save graph」機能で作成する。
PNG ファイル	Draw graph タブの「Save graph」機能で作成する。

## 5. Statistics file

G 検定による k-mer 配列の検出結果ファイルである out\_prefix.statistics.txt の内容を Table 4 に、出力例を Table 5 に示した。データはタブ区切りで出力される。1 行目には解析した K-mer の塩基数がコメント形式で出力され、2 行目にはヘッダー情報がコメント形式で出力される。

**Table 4 out\_prefix.statistics.txt の内容 (ヘッダー情報)**

ヘッダー	内容
Pos	ベクター配列上の位置
Seq	塩基
Mutant	Mutant における k-mer のカウント数
WildType	Wild type における k-mer のカウント数
Gval	G 統計量 (Williams Collections)
Pval	P 値
FDR	FDR (Benjamin-Hochberg)
Bonferroni	Bonferroni

**Table 5 out\_prefix.statistics.txt の例**

#K-mer	20						
#Pos	Seq	Mutant	WildType	Gval	Pval	FDR	Bonferroni
1	T	1	0	0.9769876	0.32294366	0.372447	1.0
2	A	1	0	0.9769876	0.32294366	0.3724235	1.0
3	A	1	0	0.9769876	0.32294366	0.37240002	1.0
4	A	1	0	0.9769876	0.32294366	0.3723765	1.0
5	C	1	0	0.9769876	0.32294366	0.37235302	1.0
1146	G	12	0	16.893818	3.9530132E-5	3.5464566E-4	0.7224132
1147	C	12	0	16.893818	3.9530132E-5	3.5447165E-4	0.7224132
1148	T	12	0	16.893818	3.9530132E-5	3.5429778E-4	0.7224132
1149	C	13	0	18.358307	1.830192E-5	2.1803624E-4	0.3344676
1150	A	13	0	18.358307	1.830192E-5	2.178942E-4	0.3344676
:	:	:	:	:	:	:	:

## 6. Outside file

検出された k-mer 配列周辺の解析結果ファイルである `out_prefix.outside.txt` の出力例を Table 6 に、各列のデータの内容を Table 7 に示した。データはタブ区切りで出力される。1 行目には、解析した k-mer の塩基数、周辺配列を解析対象とする FDR の閾値、上下流に付加した塩基数（片側当り）がコメント形式で出力される。2 行目以降には、「K-mer 配列行（グレー行）」と「両側付加塩基行（白行）」が出力される。なお、記号 A, B, C, ...そのものは出力されない。

**Table 6** `out_prefix.outside.txt` の例

A	B	C	D	E	F	G	H	I
#K-mer	20	FDR	0.01	Bases	5			
2324	2	ACATATGCCCCG TCGACCCCA	82	0	111.752	4.04896e-26	1.01017e-25	7.32376e-22
TTGAT	TCACA	72	0	TTGATACATATGCCCCG TCGACCCCA TCACA	98.0146	4.15309e-23	1.75028e-22	1.36662e-18
TTCAT	TCACA	1	0	TTCATACATATGCCCCGT CGACCCCA TCACA	0.911434	0.339734	0.431766	1
2325	3	CATATGCCCCGT CGACCCCAT	82	0	111.752	4.04896e-26	1.01003e-25	7.32376e-22
TGATA	CACAA	75	0	TGATACATATGCCCCGT CGACCCCAT CACAA	102.126	5.21069e-24	2.28131e-23	1.71463e-19
TGATA	CACAT	1	0	TGATACATATGCCCCGT CGACCCCAT CACAT	0.911434	0.339734	0.431733	1
TCATA	CACAA	1	0	TGATACATATGCCCCGT CGACCCCAT CACAA	0.911434	0.339734	0.43175	1
2326	3	ATATGCCCCGTC GACCCCATC	81	0	110.381	8.0853e-26	2.00009e-25	1.46247e-21
GATAC	ACAAG	75	0	GATACATATGCCCCGTC GACCCCATC ACAAG	102.126	5.21069e-24	2.281e-23	1.71463e-19
GATAC	ACATG	1	0	GATACATATGCCCCGTC GACCCCATC ACATG	0.911434	0.339734	0.4317	1
CATAC	ACAAG	1	0	GATACATATGCCCCGTC GACCCCATC ACAAG	0.911434	0.339734	0.431716	1
:	:	:	:	:	:	:	:	:

**Table 7 out\_prefix.outside.txt の内容**

記号	K-mer 配列行 (グレー行)	両側付加塩基行 (白行)
A	ベクター配列上の位置	上流側付加塩基
B	両側付加塩基行 (白) 行数	下流側付加塩基
C	K-mer 配列	Mutant における付加塩基のカウント数
D	Mutant における k-mer のカウント数	Wild type における付加塩基のカウント数
E	Wild type における k-mer のカウント数	K-mer 配列に上下流の塩基を付加した配列
F	G 統計量 (Williams Collections)	
G	P 値	
H	FDR (Benjamin-Hochberg)	
I	Bonferroni	

## 7. Mer frequency file

指定した k-mer 長のベクター配列上の全てのパターンに対して、Mutant および Wild type の各々の NGS データ内でのカウント数を入力する。Mer frequency file の出力例を Table 8 に示した。データはタブ区切りで出力される。

**Table 8 Mer frequency file の例**

K-mer 配列	カウント数
AAAAAAAAAGGAGAACACAT	140
AAAAAAAAAGCATGAAAAGAT	96
AAAAAAAAAGGAGAACACATG	138
AAAAAAAAAGGATGATCATGC	109
AAAAAAAAATATGTGGTAATT	122
AAAAAAAAATCATGAAATCGA	138
AAAAAACATGTCATAACAA	0
AAAAAACACCGCTACCAG	81
AAAAAACTAAAATAGAGTT	124
AAAAAACTAAGGAAACATT	108

## 8. Count mer

「Count mer」の画面を Figure 1 に、機能・操作方法を Table 9 に示した。⑦K-mer の塩基数は便宜的に 8～1024 に制限されている。⑫Maximum number of threads は PC の論理プロセッサ数と 8 の少ない方が設定される（変更可能）。実際に利用されるプロセッサ数は、②と④で指定した fastq ファイルの総数が上限になる。⑬Stop と⑭Execute はどちらか一方のボタンが有効になる。

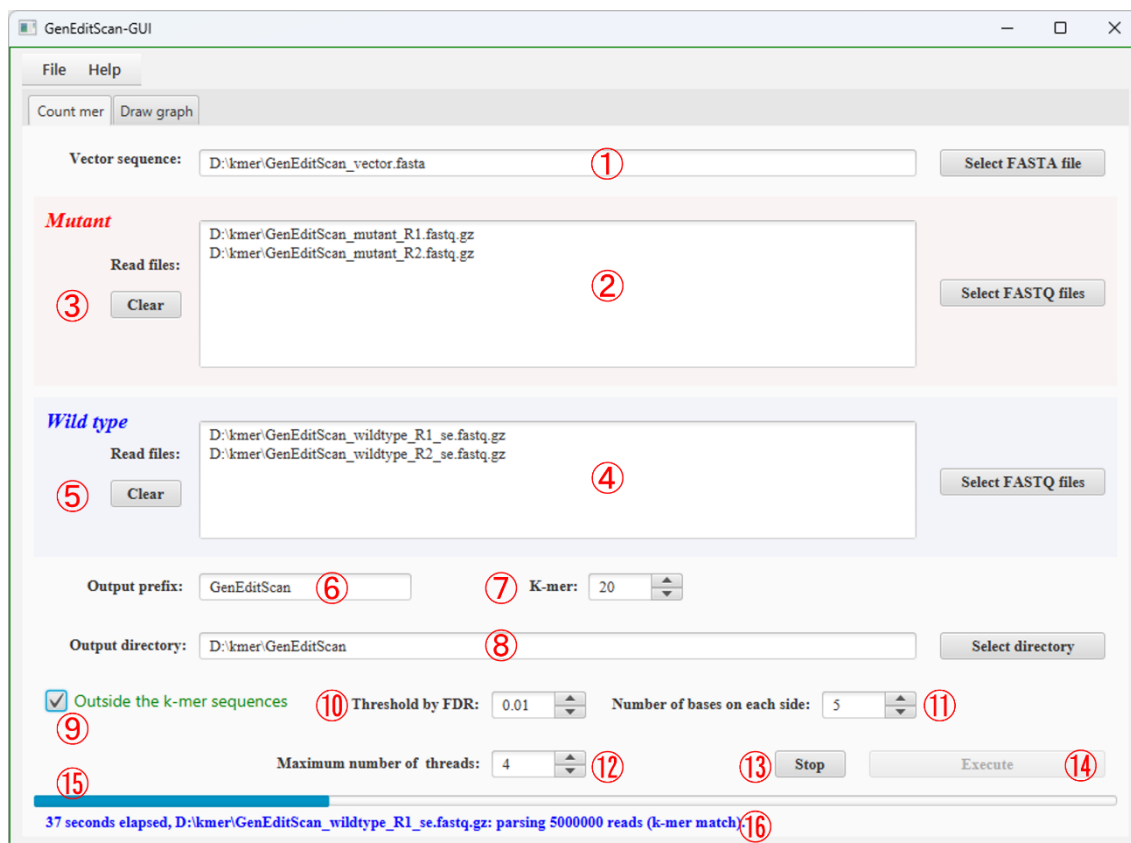


Figure 1 Count mer の画面

Table 9 Count mer の機能・操作方法

No.	表示	機能・操作方法
①	Vector sequence	右の「Select FASTA file」ボタンで選択
②	Mutant Read files	右の「Select FASTQ files」ボタンで選択
③	Clear	②をクリア
④	Wild type Read files	右の「Select FASTQ files」ボタンで選択
⑤	Clear	④をクリア
⑥	Output prefix	出力ファイルの PREFIX
⑦	K-mer	解析する K-mer（便宜上 8～1,024 に制限）

⑧	Output directory	右の「Select directory」ボタンで選択
⑨	Outside the k-mer sequences	K-mer の周辺配列の解析を実行
⑩	Threshold by FDR	周辺配列を解析する k-mer の FDR の閾値
⑪	Number of bases on each side	K-mer の両端の各々の塩基数
⑫	Maximum number of threads	最大スレッド数 (fastq ファイルの総数が実際に利用されるスレッド数の上限)
⑬	Stop	計算の中止
⑭	Execute	計算の実行
⑮	Progress bar	計算の進行状況
⑯	Message	計算状況の出力



## 9. Draw graph

### 9.1. グラフ表示機能

「Draw graph」の画面を Figure 2 に、機能・操作方法を Table 10 に示した。本画面は「Count mer」の実行により自動で描画される。計算済みの⑱Statistics file の選択による描画も可能である。⑳Save graph では、ベクトル形式の PDF ファイルおよび PNG ファイルへの保存が可能である。



Figure 2 Draw graph の画面

Table 10 Draw graph の機能・操作方法

No.	表示	機能・操作方法
①	Y-axis (upper)	Y 軸のタイトル。
②	auto	上図の Y 軸を自動スケール
③	specify	上図の Y 軸の範囲を指定
④	Mutant	上図の Mutant の line color
⑤	Wild type	上図の Wild type の line color
⑥	Y-axis (lower)	下図の Y 軸の目盛

⑦	auto	下図の Y 軸を自動スケール
⑧	specify	下図の Y 軸の範囲を指定
⑨	Significant	下図の Significant の line color
⑩	Not significant	下図の Not significant の line color
⑪	Threshold	下図の閾値の直線と line color
⑫	X-axis	X 軸のタイトル
⑬	auto	X 軸を自動スケール
⑭	specify	X 軸の範囲を指定
⑮	Rotate yticks	Y 軸目盛の回転
⑯	Save graph	グラフの保存 (PDF、PNG)
⑰	Clear	グラフ表示のクリア
⑱	Redraw	グラフの再描画
⑲	Statistics file	右の「Select statistics.txt」ボタンで選択
⑳	Outside file:	k-mer 配列とその周辺配列の情報をポップアップで表示
㉑	Outside file	右の「Select outside.txt」ボタンで選択

「⑥Y-axis (lower)」で G-statistics を選択時に、G 統計量に与える閾値⑪の例を Table 11 に示した。自由度 1 のカイ二乗分布表により値が決定される。

**Table 11 自由度 1 のカイ二乗分布表 (G 統計量)**

確率	0.050	0.025	0.010	0.005
G 統計量	3.84146	5.02389	6.63490	7.87944

「⑥Y-axis (lower)」の選択メニューにより、Y 軸の内容・タイトルを選択することができる (図 1)。

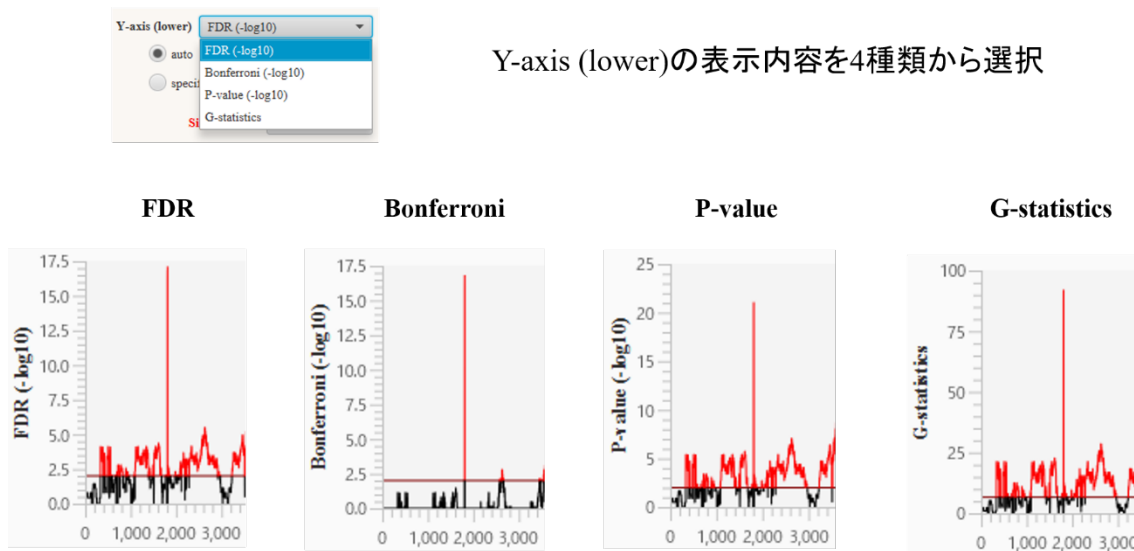


図 1 Y 軸の内容・タイトルの選択

グラフの横軸範囲を変更するには、Figure 2 の「X-axis」の「specify」を指定する方法の他に、拡大したい範囲をマウス操作により選択することで変更できる(Figure 3)。

「クリック」→「そのまま横方向に移動」→「リリース」  
(上下どちらの画面でもOK:連動)

「クリック」と「リリース」の範囲を拡大

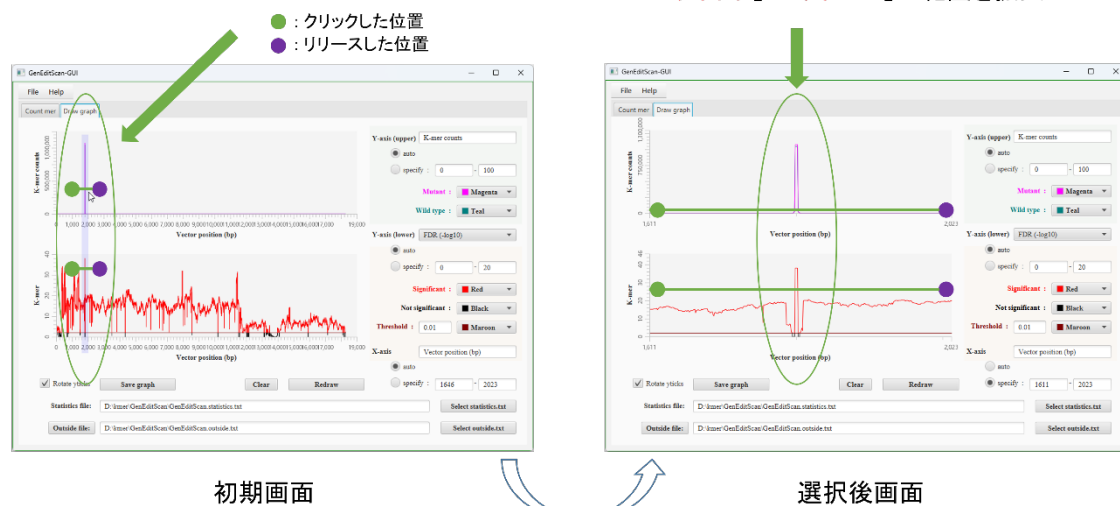


Figure 3 マウスによる横軸範囲の選択

グラフの縦軸の最大値を指定するには、Figure 2 の「Y-axis (upper)」(上の画面) または「Y-axis (lower)」(下の画面) の「specify」を指定する方法の他に、ダブルクリックをした位置を最大値に指定することが可能である(Figure 4)。

「ダブルクリック」の位置を最大値に変更  
(上下どちらの画面でもOK:独立)

● : ダブルクリックした位置

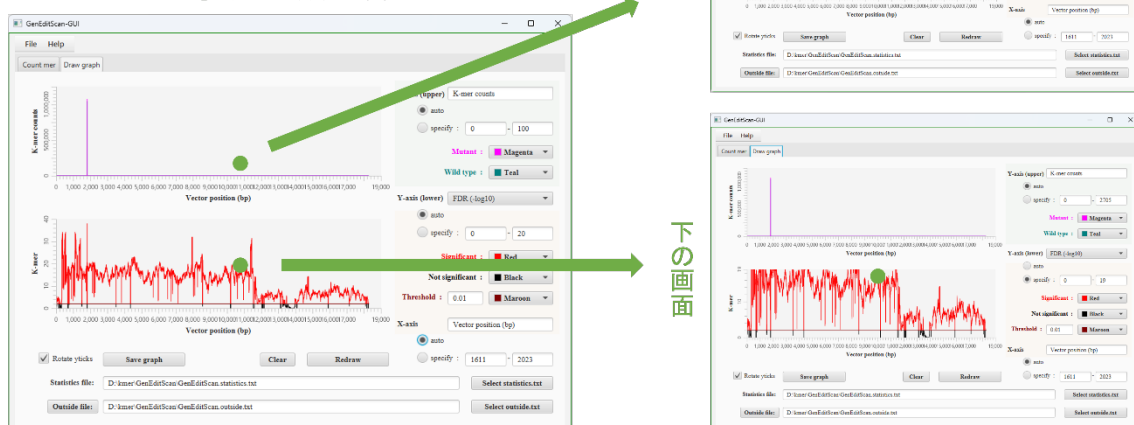


Figure 4 マウスによる縦軸の最大値の指定

Draw graph の「Select statistics.txt」ボタンでは、G 検定の結果ファイルである statistics.txt ファイルの選択が可能である (Figure 5)。

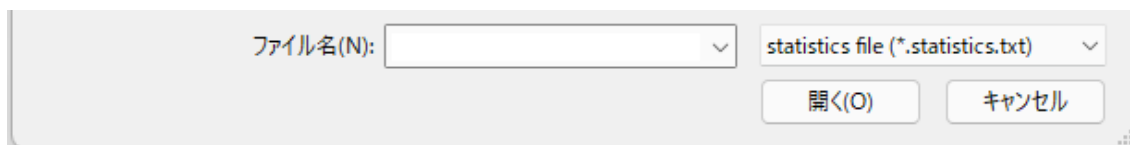


Figure 5 「Select statistics.txt」ボタンによるファイルの入力

Draw graph の「Select outside.txt」ボタンでは、検出された k-mer 配列周辺の解析結果ファイルである outside.txt ファイルの選択が可能である (Figure 6)。

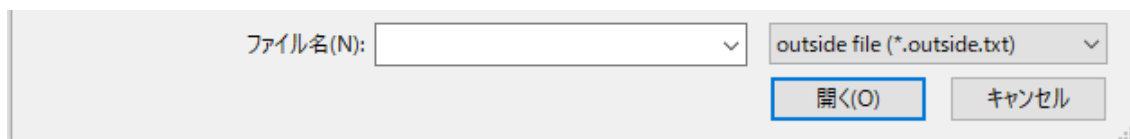


Figure 6 「Select outside.txt」ボタンによるファイルの入力

Draw graph の「Save graph」ボタンでは、グラフの保存形式を PDF と PNG から選択することが可能である (Figure 7)。

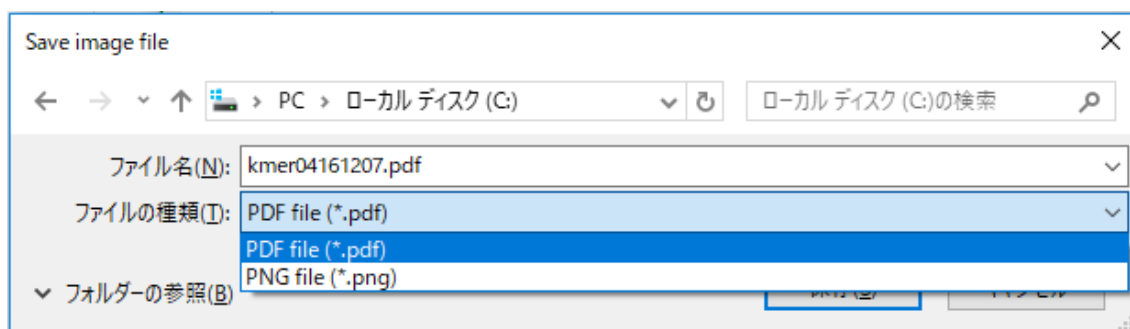
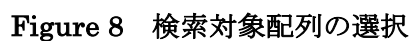


Figure 7 「Save graph」ボタンによるファイルの保存

「Draw graph」タブの下の画面で、FDR が閾値を超えている領域をクリックすることで、検出された k-mer 配列とその周辺の解析結果を表示することができる(Figure 8)。



13

検出された k-mer 配列とその周辺の解析結果の表示内容を Figure 9 に示した。各々の項目をクリックすることで昇順・降順の並べ替えが可能である。また、Sequence 列は右クリックによりコピーが可能である。

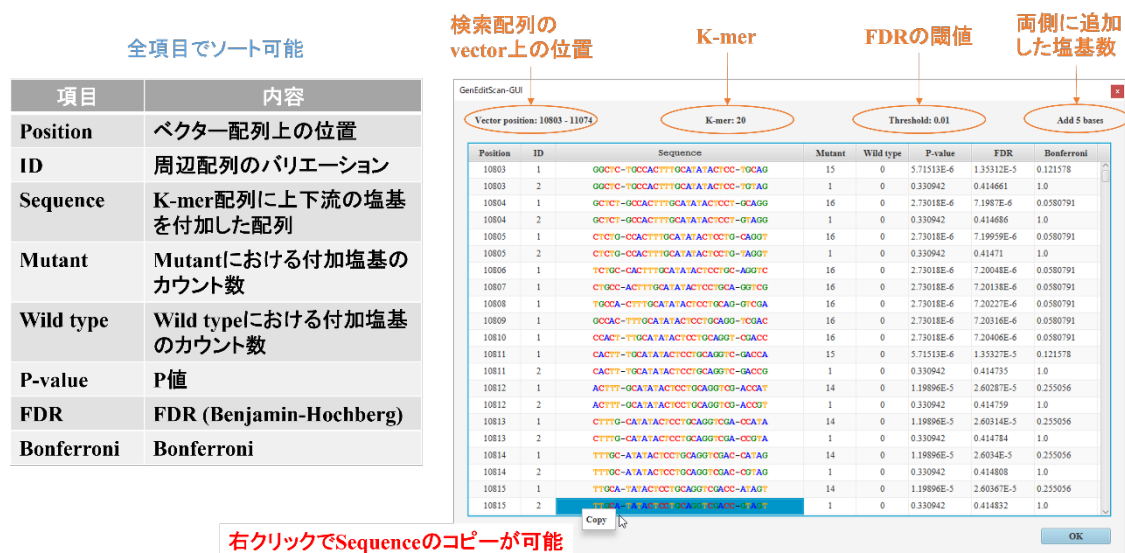


Figure 9 検出された k-mer 配列とその周辺の解析結果の表示内容

10. Menu

10.1. File

「File」の画面を Figure 10 に、機能・操作方法を Table 12 に示した。設定ファイル (configuration file)により、保存時の画面の状態を再現することが可能である。

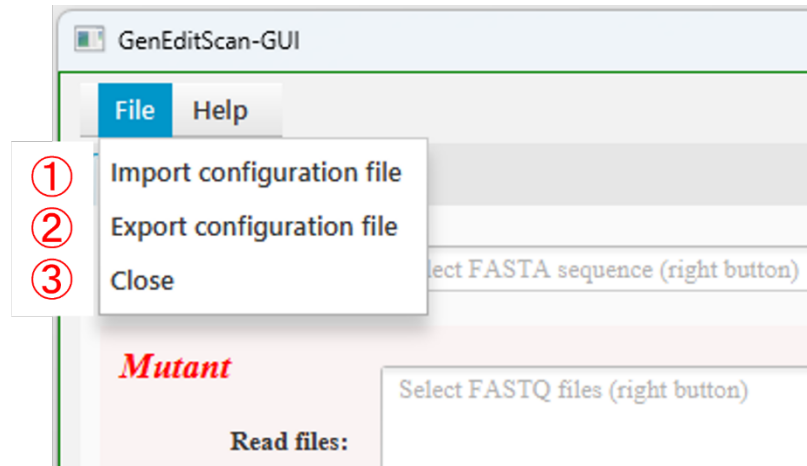


Figure 10 File の画面

Table 12 File の機能・操作方法

No.	表示	機能・操作方法
①	Import configuration file	設定ファイルの出力
②	Export configuration file	設定ファイルの入力
③	Close	プログラムの終了



10.2. Help

「Help」の画面を Figure 11 に、機能・操作方法を Table 13 に示した。利用している PC において、PDF ファイルが関連付けられているアプリケーションにより User guide が表示される。

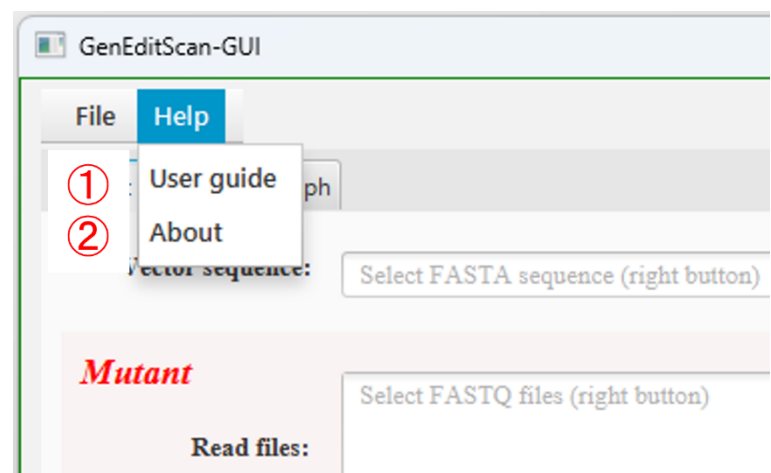


Figure 11 Help の画面

Table 13 Help の機能・操作方法

No.	表示	機能・操作方法
①	User guide	User guide の表示
②	About	プログラムのバージョン・ライセンスの表示

## 11. License

Copyright 2019-2024 National Agriculture and Food Research Organization (NARO)

## 12. Open source licenses

GenEditScan-GUI では、ベクトル形式の PDF ファイルに画像を保存するため、および、P 値の算出用に、Table 14 に示したツールを外部ライブラリとして利用している。

**Table 14 Open source licenses**

Open source	Licenses
Apache Batik	<p>Copyright 1999-2022 The Apache Software Foundation</p> <p>This product includes software developed at The Apache Software Foundation (<a href="http://www.apache.org/">http://www.apache.org/</a>).</p> <p>This software contains code from the World Wide Web Consortium (W3C) for the Document Object Model API (DOM API) and SVG Document Type Definition (DTD).</p> <p>This software contains code from the International Organisation for Standardization for the definition of character entities used in the software's documentation.</p> <p>This product includes images from the Tango Desktop Project (<a href="http://tango.freedesktop.org/">http://tango.freedesktop.org/</a>).</p> <p>This product includes images from the Pasodoble Icon Theme (<a href="http://www.jesusda.com/projects/pasodoble">http://www.jesusda.com/projects/pasodoble</a>).</p>
Apache FOP	<p>Copyright 1999-2024 The Apache Software Foundation</p> <p>This product includes software developed at The Apache Software Foundation (<a href="http://www.apache.org/">http://www.apache.org/</a>).</p>
JFXConverter	<p>Copyright (c) 2016, 2020 Herve Girod</p> <p>All rights reserved.</p>

	<p>Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:</p> <ol style="list-style-type: none"> <li>1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.</li> <li>2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.</li> </ol> <p>THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.</p> <p>The views and conclusions contained in the software and documentation are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the FreeBSD Project.</p> <p>Alternatively if you have any questions about this project, you can visit the project website at the project page on <a href="https://sourceforge.net/projects/jfxconverter/">https://sourceforge.net/projects/jfxconverter/</a></p>
--	--

Colt	<p>Copyright (c) 1999 CERN - European Organization for Nuclear Research.</p> <p>Permission to use, copy, modify, distribute and sell this software and its documentation for any purpose is hereby granted without fee, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.</p> <p>CERN makes no representations about the suitability of this software for any purpose. It is provided "as is" without expressed or implied warranty.</p>
------	---

### 13. Release notes

GenEditScan-GUI の開発履歴を Table 15 に示した。

**Table 15 Version history and release notes**

Version	Date	Release notes
1.0.2	December XX, 2024	ソースコードのリファクタリング。
1.0.0	February 29, 2024	<ul style="list-style-type: none"> <li>・ プログラム名の変更による version の振り直し。</li> <li>・ Outside file に「K-mer 配列に上下流の塩基を付加した配列」を追加入出力。この配列をコピーして DB 検索を実施することを想定。</li> <li>・ 経過時間と計算ログをプログレスバーの下に表示。</li> <li>・ 同じ値の P 値には同じ値の FDR 値を割り当てるように修正。</li> </ul>
2.1.0-beta	February 25, 2022	<ul style="list-style-type: none"> <li>・ 多重検定機能の追加。</li> <li>・ P 値、FDR、Bonferroni のグラフ表示機能の追加。</li> </ul>
1.2.0-beta	February 22, 2019	First release.