bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Analysis of Sequences of Bases of Human DNA Using the Square Wave Method (SWM)

Osvaldo Skliar* Sherry Gapper† Ricardo E. Monge‡

November 24, 2023

Abstract

A description is provided of how the Square Wave Method (SWM) can be applied to analyze sequences of bases of human DNA. The results obtained are displayed using the Square Wave Transform (SWT), an SWM tool. It is hypothesized that the results of this analysis can be of interest, in certain cases, to understand the functional role of some of those sequences. Preliminary data which make this hypothesis plausible are presented.

Keywords: Sequences of bases of human DNA, Square Wave Method (SWM), Square Wave Transform (SWT)

Mathematics Subject Classification 2020: 40-08, 92-08, 65F99, 94A12

*Universidad Nacional, Costa Rica. E-mail: osvaldoskliar@gmail.com. https://orcid.org/0000-0002-8321-3858.

†Universidad Nacional, Costa Rica. E-mail: sherry.gapper.morrow@una.ac.cr. https://orcid.org/0000-0003-4920-6977.

‡Universidad CENFOTEC, Costa Rica. E-mail: rmonge@ucenfotec.ac.cr. https://orcid.org/0000-0002-4321-5410.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1 Introduction

The Square Wave Method (SWM) is a new algebraic method for the analysis of signals and images [1], [2], [3], [4], and [5]. The Square Wave Transform (SWT) [3] is a tool belonging to the SWM which makes it possible to easily visualize the results obtained when using the method.

Mention is made here of contributions of authors other than those who introduced the SWM, in which reference is made to that method. One of them [6] states: "The idea behind our approach builds upon existing classical results that use *exact* square waves to approximate a function (39)". The only reference to those "existing classical results" is (39), which is the same as [5] in the present article. Reference to that article is also made in [7].

Other sources, [8], [9], [10],[11] and [12], also refer to that article. The SWM has been mentioned frequently in patents: [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129] and [130].

The objectives of this article are the following:

I. specify how sequences of bases of DNA can be analyzed using the SWM and the results thus obtained, presented using the SWT; and

II. formulate the hypothesis that at least in certain cases considering the sequences of bases of the nucleic acids – of DNA in particular – as signals susceptible to being analyzed with a method characteristic of signal analysis, such as the SWM, can be useful in genomics. Preliminary data which make this hypothesis plausible are presented.

2 Brief Review of the Use of the SWM for the Analysis of Numerical Sequences

As seen in the following section of this article, the first operation to be carried out to analyze a sequence of bases of DNA with the SWM is the codification of that sequence as a numerical sequence. Therefore, this section will be useful for those who are not familiar with articles on the SWM.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The SWM makes it possible to approximate numerical sequences using a certain sum, to be specified, of parts of trains of digitalized square waves. The concept of "digitalized square wave" is justified because the wave is not treated as a function with a sole discontinuity (in the transition from positive semiwave to negative semiwave, or vice versa) but rather as a sequence of numerical values.

Another concept which should be characterized in the treatment of numerical sequences using the SWM is that of frequency.

Usually, in the case of a phenomenon that is repeated with the same characteristics, by *frequency* of that phenomenon one understands the number – n – of occasions that it occurs in each unit of time. If the second – s – is taken as the unit of time, then the *frequency* of this phenomenon is defined as: $\frac{n}{s} = ns^{-1} = n$ Hz. When analyzing a numerical sequence using the SWM, the unit of time is replaced by a "numerical unit" composed of as many elemental numerical subunits as the number of numerical values that the sequence has. Thus, for example, consider the following numerical sequence:

$$S = [75, -54, 38, 125, -67, 81, -90, 49].$$

The above numerical sequence is composed of 8 elemental subunits which are the numerical values that constitute that sequence.

The elements corresponding to the phenomena that are repeated with the same characteristics are square waves – the elements comprising each train of square waves. Consider a train of square waves symbolized as follows:

$$\ldots A_7, A_7, -A_7, -A_7, A_7, A_7 - A_7 - A_7, \ldots$$

(A justification will be provided below for the numerical values of the subscripts used in this section for the symbols A_i for $i = 1, 2, 3, \ldots, 8$.) It can be admitted that this train is prolonged to infinity in both possible directions. Each square wave composing that train of square waves can be symbolized as $A_7, A_7, -A_7, -A_7$. The frequency corresponding to that train of square waves regarding the numerical unit specified above can be computed by dividing the number of elemental numerical subunits constituting that unit (8) by the number of elemental numerical subunits constituting the square wave considered (4): $f_7 = \frac{8}{4} = 2$. The numerical unit of reference could be thought to "cover" 2 square waves of the type considered; in other words, the event which is repeated – the square wave – occurs twice in the numerical unit of reference.

Consider another train of square waves:

$$\ldots A_8, -A_8, A_8, -A_8, A_8, -A_8, \ldots$$

In this case, the square wave constituting the above train of square waves can

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

be symbolized as $A_8, -A_8$. The frequency corresponding to that train of square waves with respect to the numerical unit of reference considered is computed as: $f_8 = \frac{8}{2} = 4$. In this case, the numerical unit of reference (composed of 8 elemental numerical subunits) "covers" 4 square waves of the type considered. In other words, 4 square waves of this type "occupy a place" in the numerical unit of reference.

Consider another train of square waves:

$$\ldots A_6, A_6, A_6, -A_6, -A_6, -A_6, A_6, A_6, A_6, -A_6, -A_6, -A_6, \ldots$$

In this case, the square wave repeated in the train of square waves specified is the following: $A_6, A_6, A_6, -A_6, -A_6, -A_6$. This square wave is composed of 6 elemental units. The frequency of the last train of square waves regarding the numerical unit of reference is: $f_6 = \frac{8}{6} = \frac{4}{3}$. Here the numerical unit of reference, composed of 8 elemental numerical subunits, "covers" only one square wave and a fraction of a square wave of the same type.

A matrix is presented below whose rows represent certain intervals (or parts) of trains of square waves used for the analysis of the numerical sequence specified. This sequence is indicated above the matrix:

$$\begin{array}{|c|c|c|c|c|c|c|c|l|}
\hline
75 & -54 & 38 & 125 & -67 & 81 & -90 & 49 & \\
\hline
A_1 & A_1 & A_1 & A_1 & A_1 & A_1 & A_1 & A_1 & f_1 = \frac{8}{16} = \frac{1}{2} \\
\hline
A_2 & A_2 & A_2 & A_2 & A_2 & A_2 & A_2 & -A_2 & f_2 = \frac{8}{14} = \frac{4}{7} \\
\hline
A_3 & A_3 & A_3 & A_3 & A_3 & A_3 & -A_3 & -A_3 & f_3 = \frac{8}{12} = \frac{2}{3} \\
\hline
A_4 & A_4 & A_4 & A_4 & A_4 & -A_4 & -A_4 & -A_4 & f_4 = \frac{8}{10} = \frac{4}{5} \\
\hline
A_5 & A_5 & A_5 & A_5 & -A_5 & -A_5 & -A_5 & -A_5 & f_5 = \frac{8}{8} = 1 \\
\hline
A_6 & A_6 & A_6 & -A_6 & -A_6 & -A_6 & A_6 & A_6 & f_6 = \frac{8}{6} = \frac{4}{3} \\
\hline
A_7 & A_7 & -A_7 & -A_7 & A_7 & A_7 & -A_7 & -A_7 & f_7 = \frac{8}{4} = 2 \\
\hline
A_8 & -A_8 & A_8 & -A_8 & A_8 & -A_8 & A_8 & -A_8 & f_8 = \frac{8}{2} = 4 \\
\hline
\end{array} \tag{1}$$

At the right of the matrix represented in (1), the corresponding frequencies for the different trains of square waves have been specified for the case considered. It can be seen that subscript 1 is used in the symbols corresponding to the numerical values of the first train of square waves and in the symbol of the

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

corresponding frequency. Subscript 2 is used in the symbols corresponding to the numerical values of the second train of square waves and in the symbol of the respective frequency; and so on, until finally, subscript 8 is used in the symbols of the numerical values of the eighth train of square waves and in the symbol of the respective frequency. From each of these 8 trains of square waves, only 8 elemental subunits are given in (1), that is, the same number of elemental units that comprise the numerical sequence to be analyzed.

In the first row of the matrix presented in (1), one semiwave of the square wave has been symbolized. The semiwave not symbolized is composed of another 8 symbols: $-A_1, -A_1, -A_1, -A_1, -A_1, -A_1, -A_1, -A_1$. Each of the square waves that are constituents of the first train of square waves is composed of 16 elemental numerical subunits. In other words, the length of wave λ_1 of each square wave in that train can be considered equal to 16: $\lambda_1 = 16$. As indicated above, $f_1 = \frac{8}{16} = \frac{1}{2}$.

Note that the number of elemental subunits of each of the square waves comprising the train of square waves corresponding to the second row of the matrix represented in (1) is 14; that is, there are 2 elemental subunits less than in each square wave of the first train of square waves. Given that $\lambda_2 = 14$, $f_2 = \frac{8}{14} = \frac{4}{7}$. The number of elemental subunits of each square wave of the third train of square waves – the one corresponding to the third row in the matrix represented in (1) – is 12; or that is, 2 subunits less in each square wave than in the second train of square waves. Thus, $f_3 = \frac{8}{12} = \frac{2}{3}$, and so on. Therefore, the following results are obtained: $f_4 = \frac{8}{10} = \frac{4}{5}$, $f_5 = \frac{8}{8} = 1$, $f_6 = \frac{8}{6} = \frac{4}{3}$, $f_7 = \frac{8}{4} = 2$, and finally $f_8 = \frac{8}{2} = 4$. The above results are given on the right of the matrix presented in (1).

To obtain the system of linear algebraic equations whose solution makes it possible to compute the numerical values of $A_1, A_2, A_3, A_4, A_5, A_6, A_7$, and A_8, the following procedure is carried out. The equality is established of a) the sum of all the elements in the first column of the matrix in (1) and b) the numerical value indicated immediately above that column: 75. This numerical value is the first in the numerical sequence to be analyzed. Hence, the following equation is obtained:

$$A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 = 75$$

The equality is established below of a) the sum of all the elements composing the second column of the matrix represented in (1), and b) the numerical value indicated immediately above that column: -54. This numerical value is the second in the numerical sequence analyzed. Thus the following equation is obtained:

$$A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 - A_8 = -54$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The same procedure is carried out with each of the following columns of the matrix represented in (1): For each of the columns the equality is established of a) the sum of all the elements that compose it, and b) the numerical value indicated immediately above the respective column. Thus the following system of linear algebraic equations is obtained:

$$\begin{array}{rcr}
A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 & = & 75 \\
A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 - A_8 & = & -54 \\
A_1 + A_2 + A_3 + A_4 + A_5 + A_6 - A_7 + A_8 & = & 38 \\
A_1 + A_2 + A_3 + A_4 + A_5 - A_6 - A_7 - A_8 & = & 125 \\
A_1 + A_2 + A_3 + A_4 - A_5 - A_6 + A_7 + A_8 & = & -67 \\
A_1 + A_2 + A_3 + A_4 + A_5 - A_6 + A_7 - A_8 & = & 81 \\
A_1 + A_2 - A_3 - A_4 - A_5 + A_6 - A_7 + A_8 & = & -90 \\
A_1 - A_2 - A_3 - A_4 - A_5 + A_6 - A_7 - A_8 & = & 49
\end{array} \tag{2}$$

When solving the above system of algebraic linear equations (2), the following numerical values are obtained for the unknowns.

$$\begin{array}{rcr@{\qquad}rcr}
A_1 & = & 170 & A_5 & = & 179 \\
A_2 & = & -134 & A_6 & = & -108 \\
A_3 & = & 23.5 & A_7 & = & 18.5 \\
A_4 & = & -138.5 & A_8 & = & 64.5
\end{array}$$

If in (2) the unknowns are replaced by the numerical values computed for them, it can be seen that the diverse equations of the system of equations (2) are satisfied.

If the order of the numerical values is inverted, a different numerical sequence is obtained which will be known as "inverted numerical sequence" with respect to the first sequence:

$$S_{inv} = [49, -90, 81, -67, 125, 38, -54, 75].$$

When applying the SWM to the analysis of this last numerical sequence, another sequence of linear algebraic equations is obtained. If the unknowns are computed in this last system of equations, the following numerical values are obtained for them:

$$\begin{array}{rcr@{\qquad}rcr}
A_1 & = & 57.5 & A_5 & = & -42.5 \\
A_2 & = & -134 & A_6 & = & 4.5 \\
A_3 & = & 136 & A_7 & = & -16 \\
A_4 & = & -26 & A_8 & = & 69.5
\end{array}$$

If in the last system of linear algebraic equations mentioned the unknowns are replaced by the numerical values computed for them, it can be seen that all of the equations in that system of equations are also satisfied.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

When applying the SWM to the analysis of a numerical sequence, there is an important dependence of the results obtained on the order of the numerical values of the sequence considered. Because the first operation to be carried out when analyzing a sequence of bases of DNA will be to assign each base a numerical value, and given that the order of the bases is very relevant for the functional role of each sequence of bases of DNA considered, this method appears adequate for the analysis of that type of sequences.

The Square Wave Transform (SWT) – a tool of the SWM – makes it possible to present, in a system of orthogonal Cartesian coordinates, the results obtained when analyzing any numerical sequence with that method. In effect, on the x-axis it is possible to plot the numerical values of the different frequencies f_i, for $i = 1, 2, 3, \ldots, n$, corresponding to the different trains of square waves, and on the y-axis, the numerical values of the different A_1, using dots or vertical lines.

In figure 1a, by using the SWT, the results are displayed of the analysis with the SWM of the first numerical sequence considered in this section.

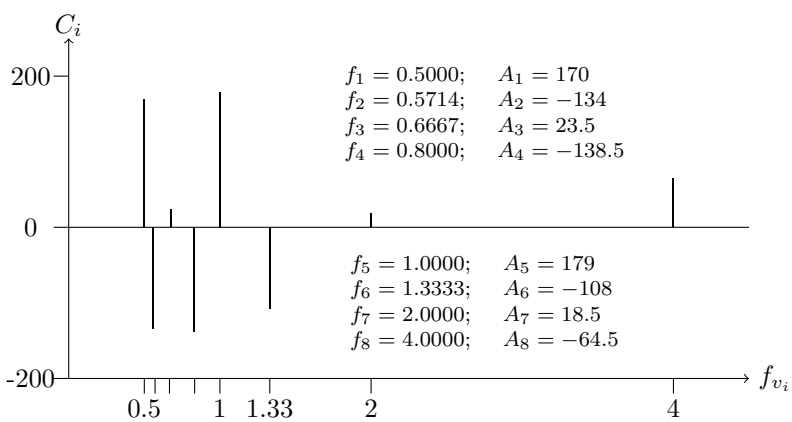

Figure 1a: SWT which presents the results of the analysis with the SWM of the first numerical sequence considered in this section

In figure 1b, by using the SWT, the results are displayed of the analysis with the SWM of the second numerical sequence considered in this section – that is, the inverted numerical sequence of the sequence previously considered.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

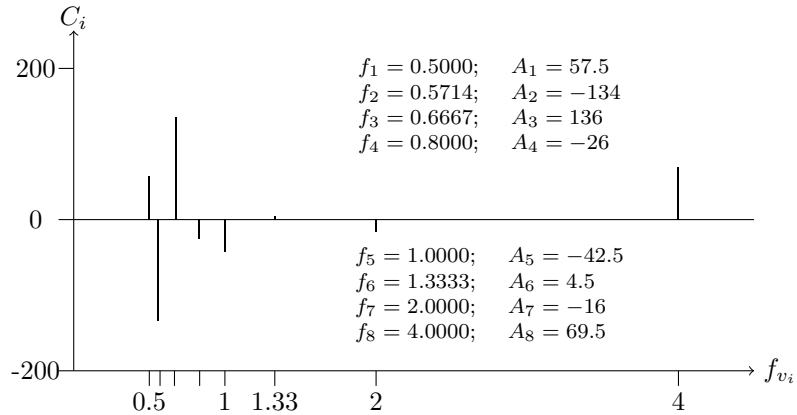

Figure 1b: SWT which presents the results of the analysis with the SWM of the inverted numerical sequence of the numerical sequence previously considered

3 Presentation—Using the SWT—of the Results of the Analysis with the SWM, of Diverse Sequences of 1200 bases of Human DNA

In this section, consideration will be given to sequences of 1200 bases of human DNA.

To use the SWM to analyze diverse sequences of these bases, the first operation to be carried out is to assign numerical values to the different bases constituting those sequences. Researchers who use the SWM to analyze sequences of bases of DNA with that objective must assign the same numerical values to those bases so that their results will be easily comparable. The assignment of numerical values to the bases of DNA used in this article is the following:

Adenine (A):	132	Guanine (G):	84
Thymine (T):	-132	Cytosine (C):	-84

In these cases, given that the interval of reference is composed of 1200 elemental subunits, the first train of square waves considered has square waves of 2400 elemental subunits. Only half a square wave of this type "fills" that interval. Therefore, $f_1 = \frac{1200}{2400} = \frac{1}{2}$.

Each square wave of the second train of square waves considered has 2 elemental subunits less than each square wave in the first train of square waves. Thus, $f_2 = \frac{1200}{2400-2} = \frac{1200}{2398}$.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The third train of square waves has the following frequency: $f_3 = \frac{1200}{2398-2}$; and so on, until finally, $f_{1199} = \frac{1200}{4} = 300$, and $f_{1200} = \frac{1200}{2} = 600$. Of course, this last result specifies that for the train of square waves with the greatest frequency, 600 square waves, each of which is composed of 2 elemental subunits, "fit" in the interval of reference composed of 1200 elemental subunits. In other words, the "repetitive event" – the square wave – "occurs" 600 times in the interval of reference: $f_{1200} = \frac{1200}{2} = 600$.

Before presenting the figures corresponding to the results of the analyses of diverse sequences of 1200 bases of human DNA, certain terminological items should be clarified.

Every square wave of the type considered is composed of 2 semiwaves: one positive and the other negative. Suppose that a numerical value of 125 corresponds to the positive semiwave. In this case, the number -125 corresponds to the negative semiwave. It is accepted that the amplitude of that square wave is equal to the absolute value of -125: $|125| = |-125| = 125$. It is also acceptable to state that the amplitude of the train of square waves composed of a sequence of that type of square waves is equal to 125.

The unknowns of the systems of linear algebraic equations which should be solved are denominated A_i, for $i = 1, 2, 3, \ldots, 1200$. Regardless of whichever of these A_i turns out to be a positive number or a negative number, the amplitude of each square wave of the ith train of square waves is $|A_i|$, the modulus of the numerical value of $|A_i|$.

For clarity, in the graphs presented in this section using the SWT, the results will be displayed only partially. In effect, only the A_i and the f_i corresponding to the 10 greatest amplitudes of all of those A_i, for $i = 1, 2, 3, \ldots, 1200$, will be shown. (Of course, there is no problem in obtaining the complete results.)

For those A_i and f_i which will be displayed in the graphs in this section, they will be denominated "prominent A_i" and "prominent f_i" respectively.

In the caption of each figure in this section, beginning with 2a and 2b, inclusive, the following will be indicated: 1) the source of the information of the data analyzed with the SWM and graphed with the SWT; 2) the number of chromosome to which the sequence of bases considered belongs; 3) the characterization of the chromosomic region considered; and 4) the specification of the sequence of bases analyzed.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

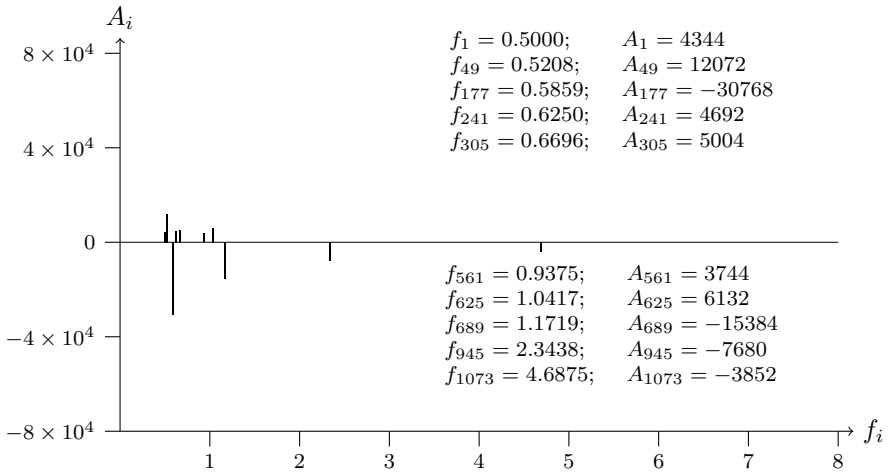

Figure 2a: T2T-CHM13v2.0 chromosome 1; p telomere; analyzed data from 1 to 1200

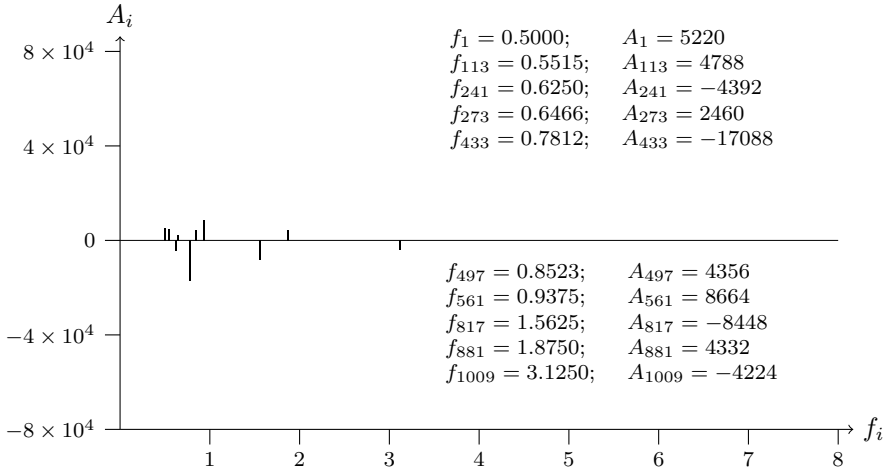

Figure 2b: T2T-CHM13v2.0 chromosome 1; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

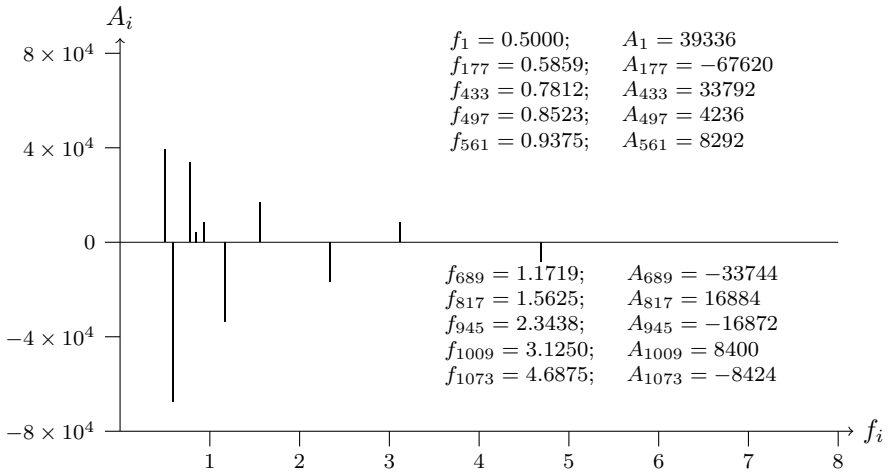

Figure 3a: T2T-CHM13v2.0 chromosome 2; p telomere; analyzed data from 1 to 1200

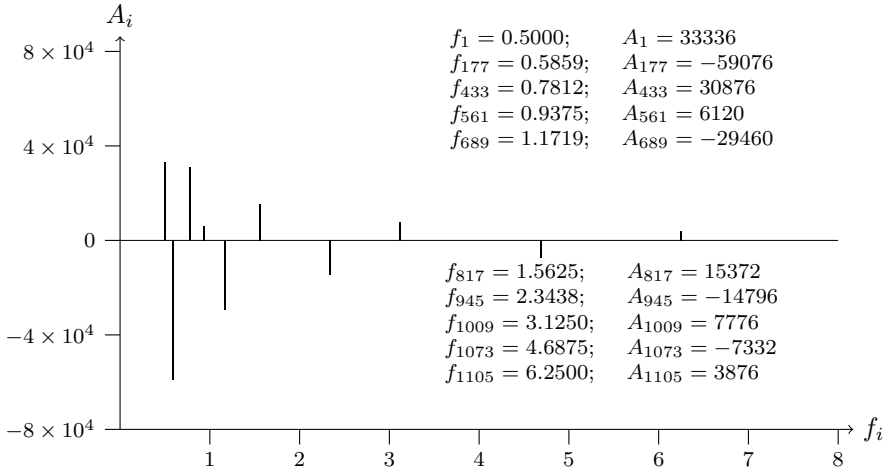

Figure 3b: T2T-CHM13v2.0 chromosome 2; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

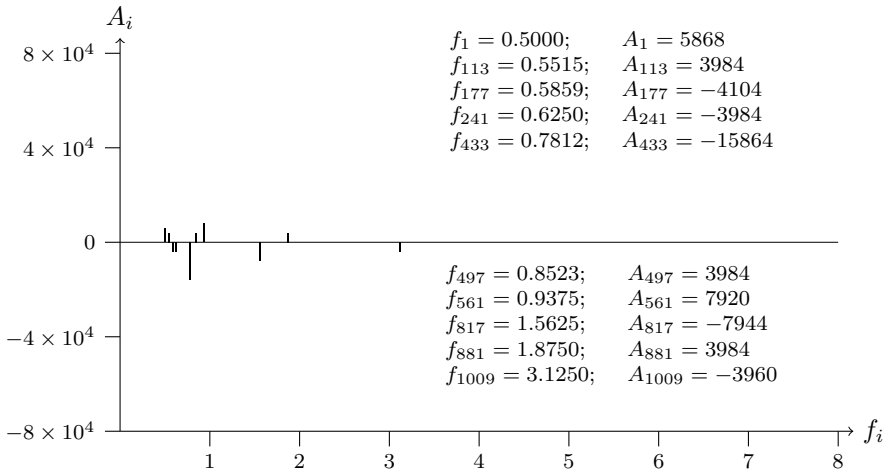

Figure 4a: T2T-CHM13v2.0 chromosome 3; p telomere; analyzed data from 1 to 1200

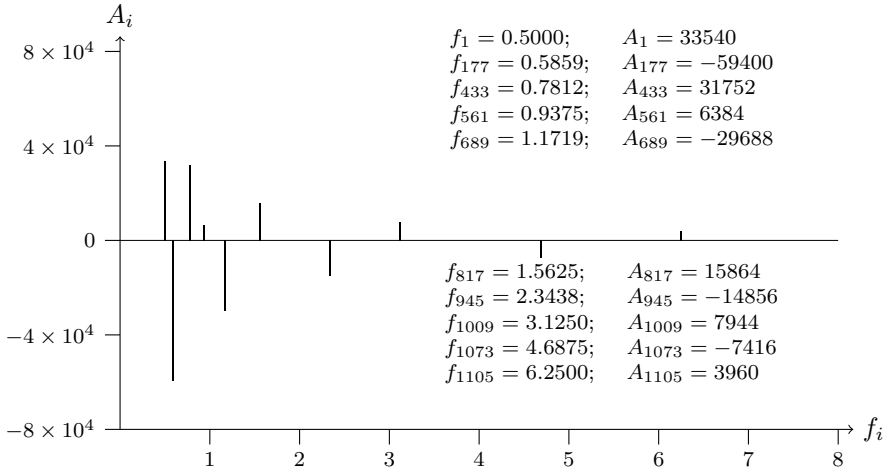

Figure 4b: T2T-CHM13v2.0 chromosome 3; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

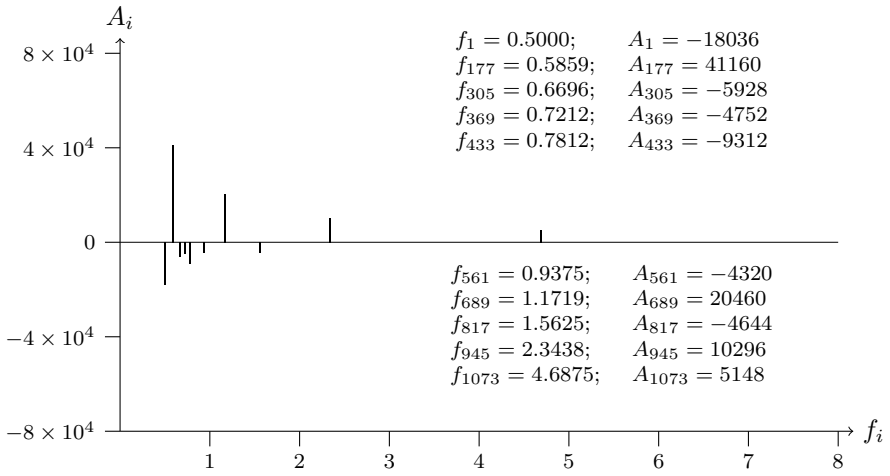

Figure 5a: T2T-CHM13v2.0 chromosome 4; p telomere; analyzed data from 1 to 1200

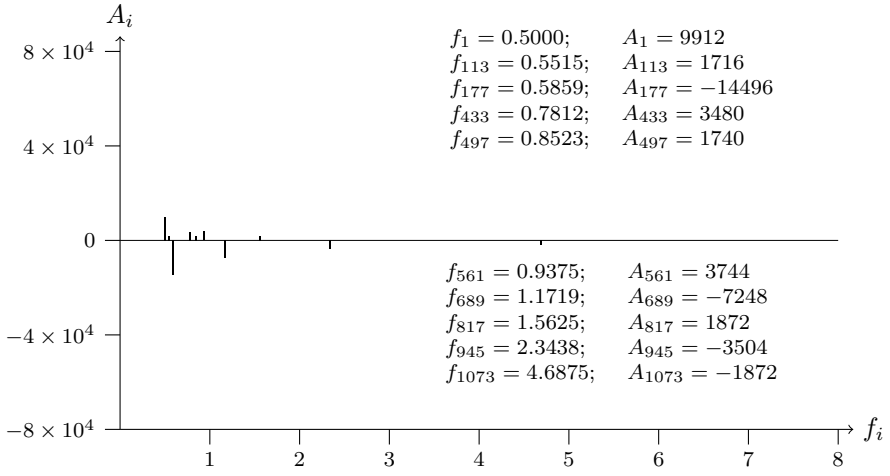

Figure 5b: T2T-CHM13v2.0 chromosome 4; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

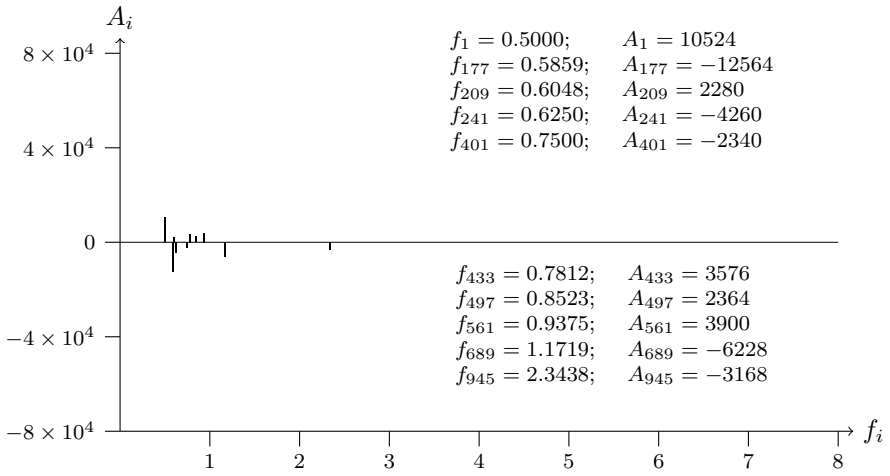

Figure 6a: T2T-CHM13v2.0 chromosome 5; p telomere; analyzed data from 1 to 1200

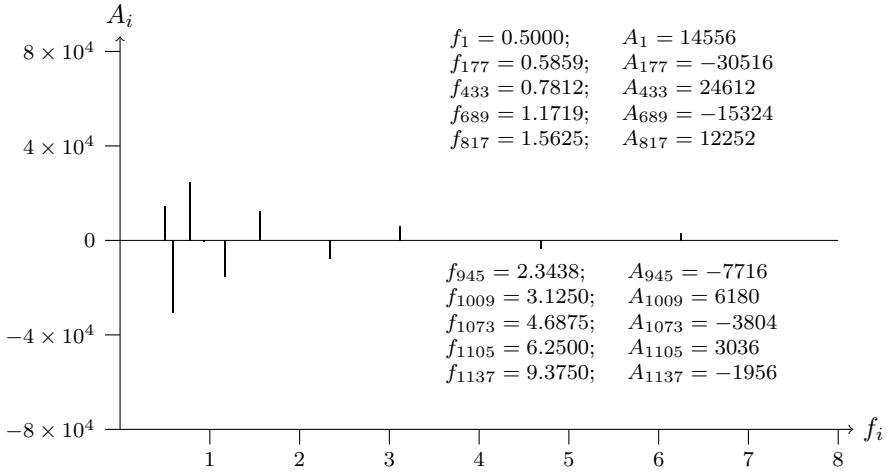

Figure 6b: T2T-CHM13v2.0 chromosome 5; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

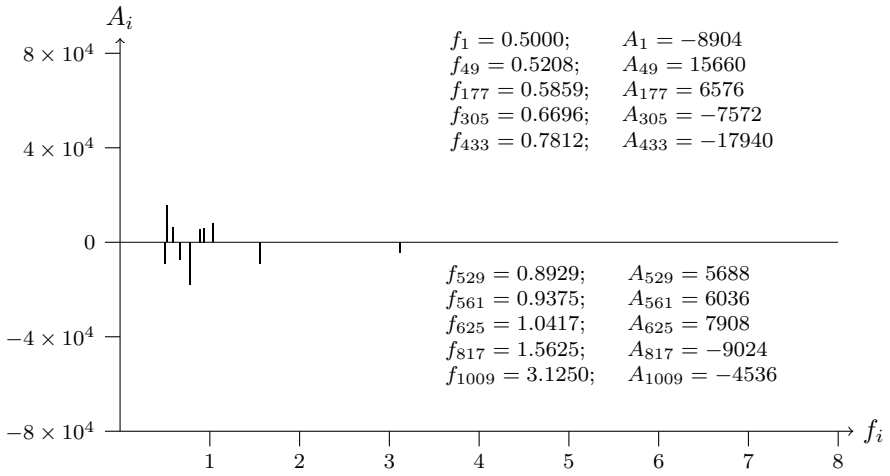

Figure 7a: T2T-CHM13v2.0 chromosome 6; p telomere; analyzed data from 1 to 1200

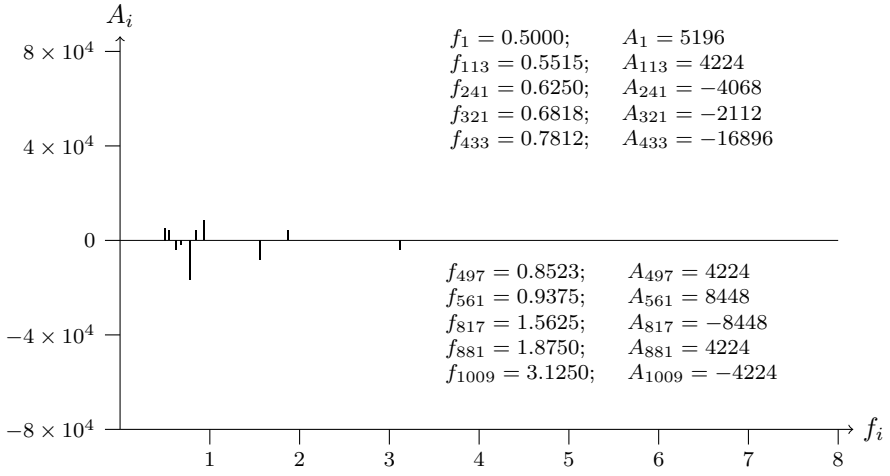

Figure 7b: T2T-CHM13v2.0 chromosome 6; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

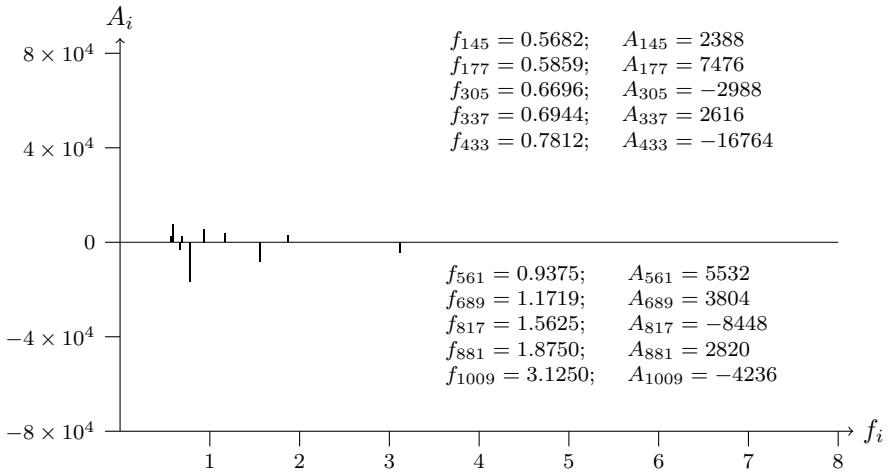

Figure 8a: T2T-CHM13v2.0 chromosome 7; p telomere; analyzed data from 1 to 1200

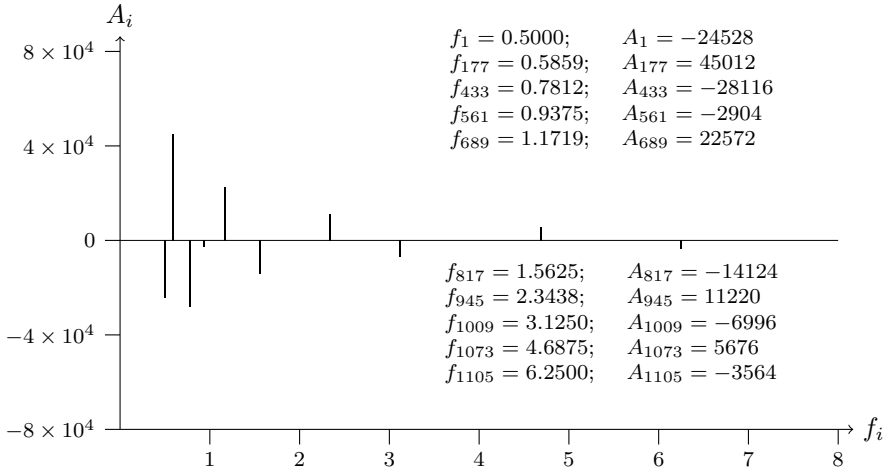

Figure 8b: T2T-CHM13v2.0 chromosome 7; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

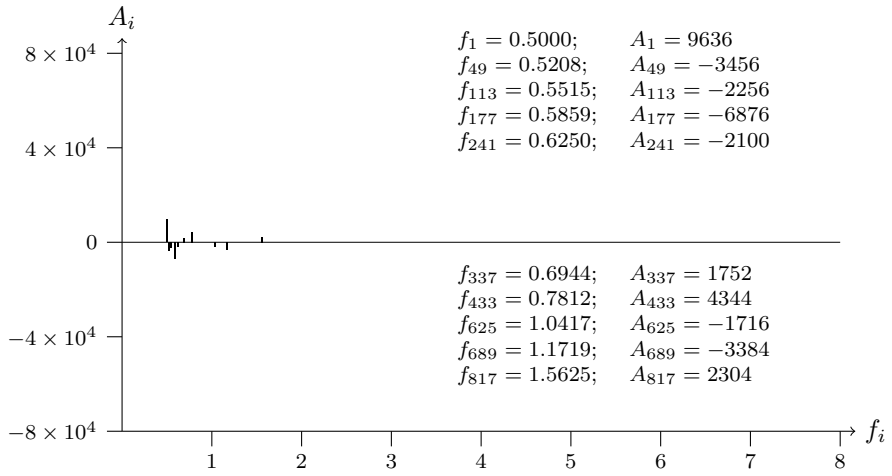

Figure 9a: T2T-CHM13v2.0 chromosome 8; p telomere; analyzed data from 1 to 1200

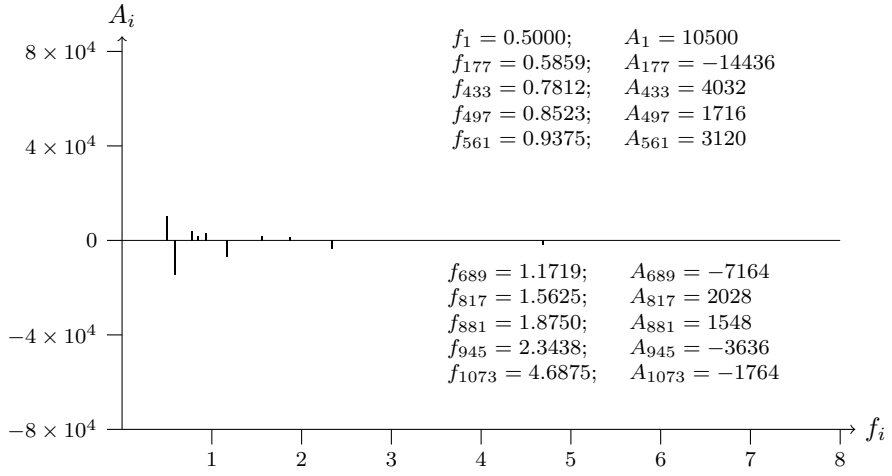

Figure 9b: T2T-CHM13v2.0 chromosome 8; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

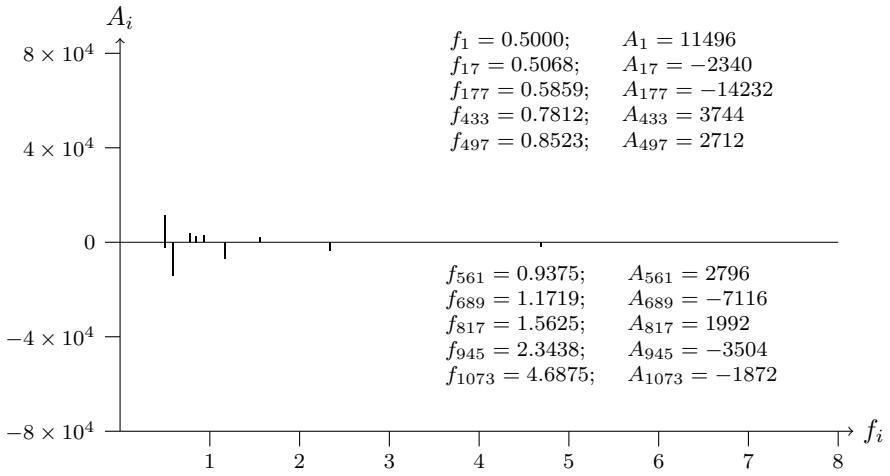

Figure 10a: T2T-CHM13v2.0 chromosome 9; p telomere; analyzed data from 1 to 1200

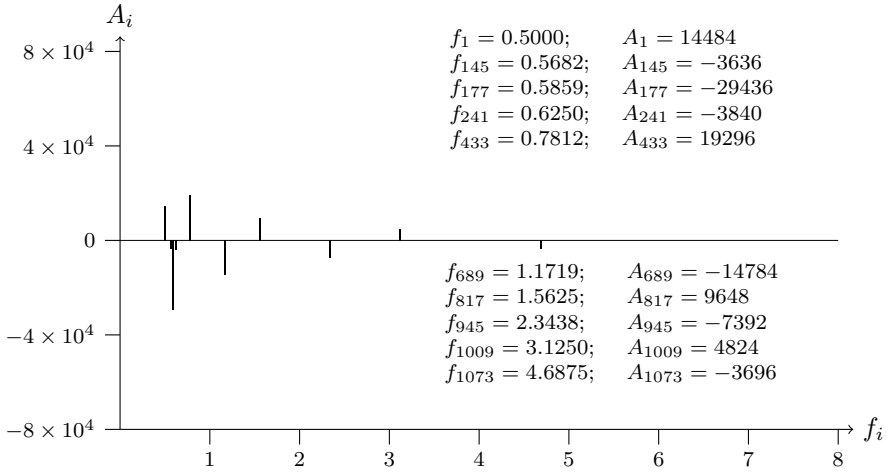

Figure 10b: T2T-CHM13v2.0 chromosome 9; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

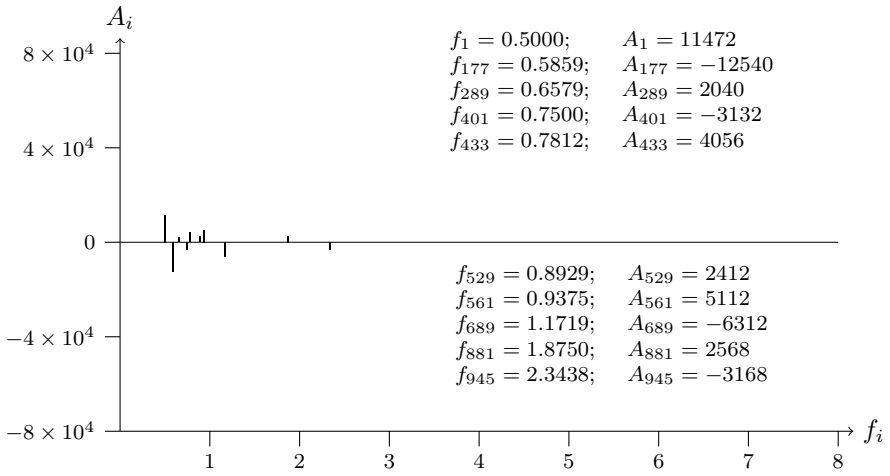

Figure 11a: T2T-CHM13v2.0 chromosome 10; p telomere; analyzed data from 1 to 1200

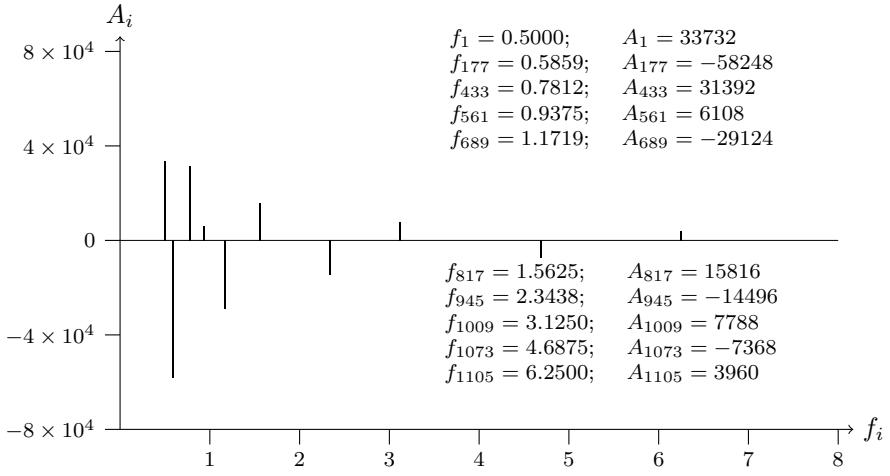

Figure 11b: T2T-CHM13v2.0 chromosome 10; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

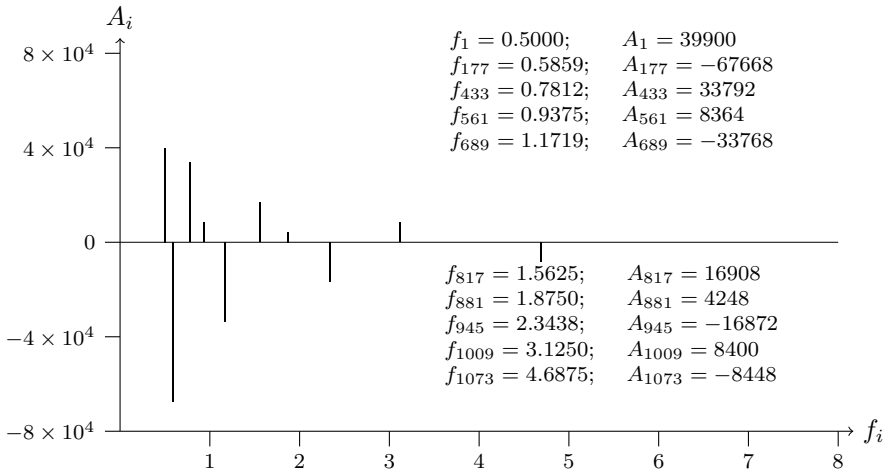

Figure 12a: T2T-CHM13v2.0 chromosome 11; p telomere; analyzed data from 1 to 1200

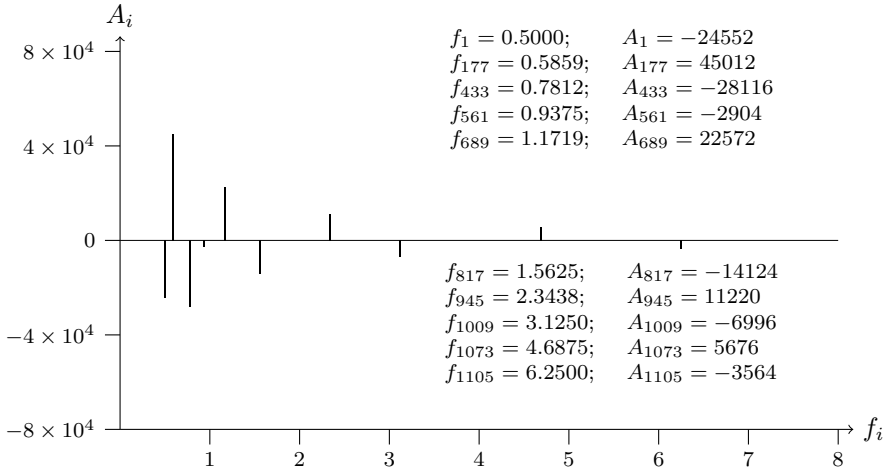

Figure 12b: T2T-CHM13v2.0 chromosome 11; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

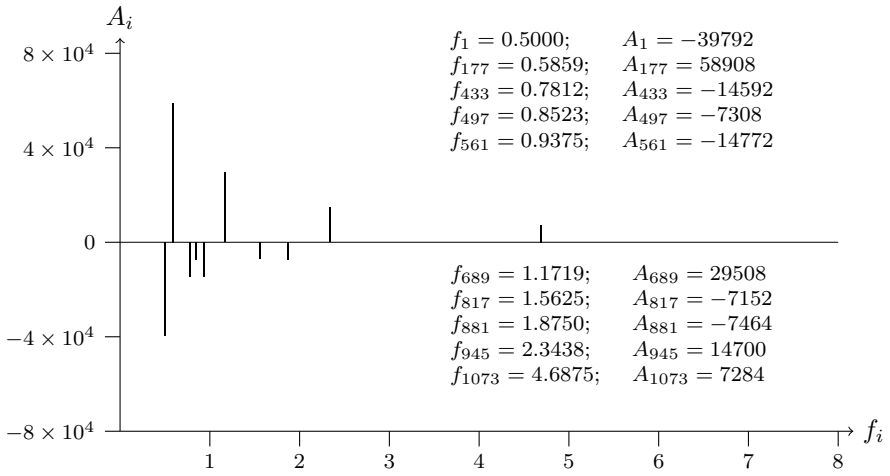

Figure 13a: T2T-CHM13v2.0 chromosome 12; p telomere; analyzed data from 1 to 1200

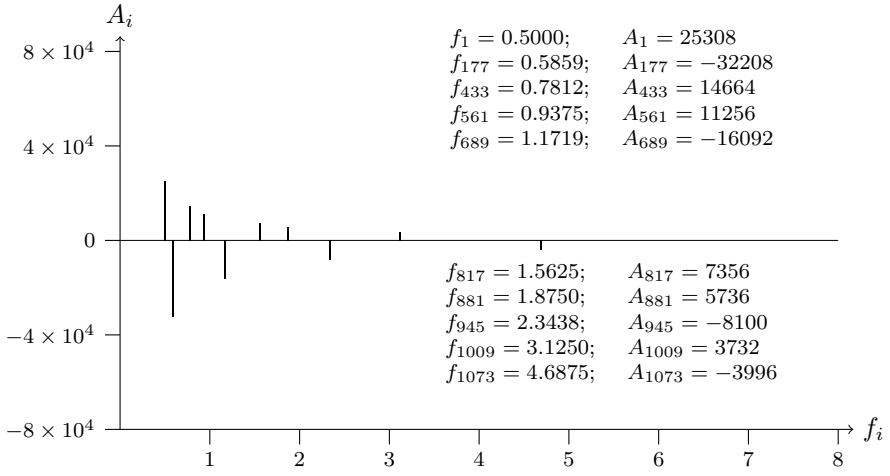

Figure 13b: T2T-CHM13v2.0 chromosome 12; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

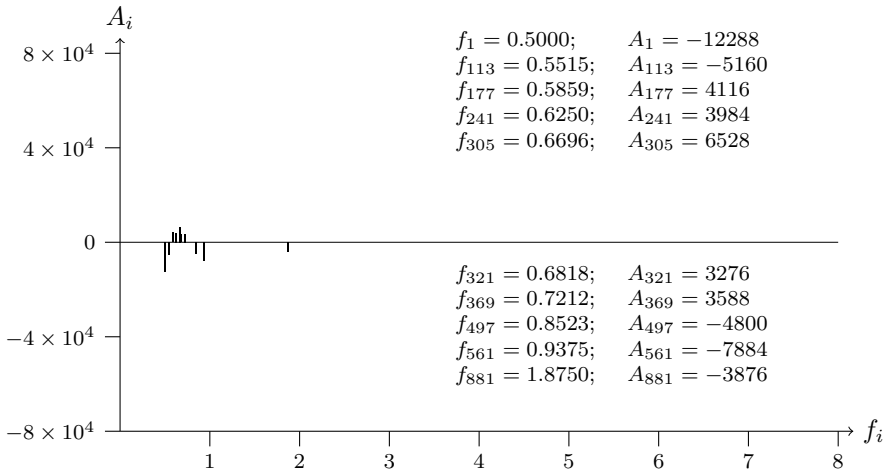

Figure 14a: T2T-CHM13v2.0 chromosome 13; p telomere; analyzed data from 1 to 1200

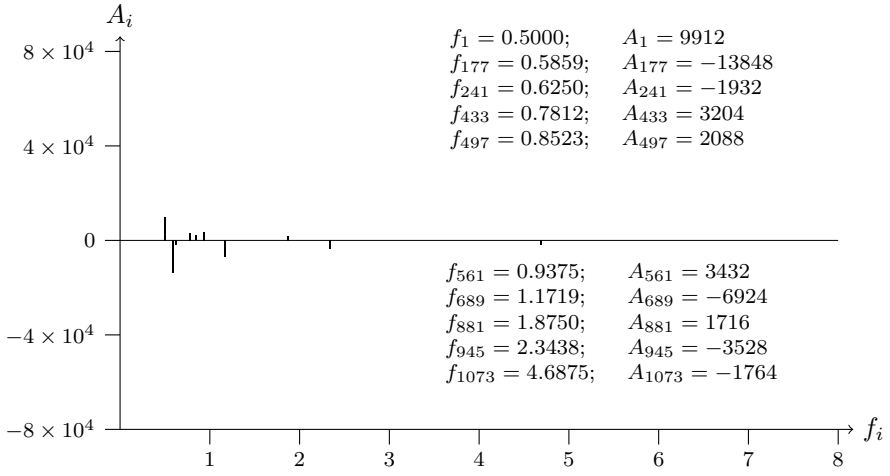

Figure 14b: T2T-CHM13v2.0 chromosome 13; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

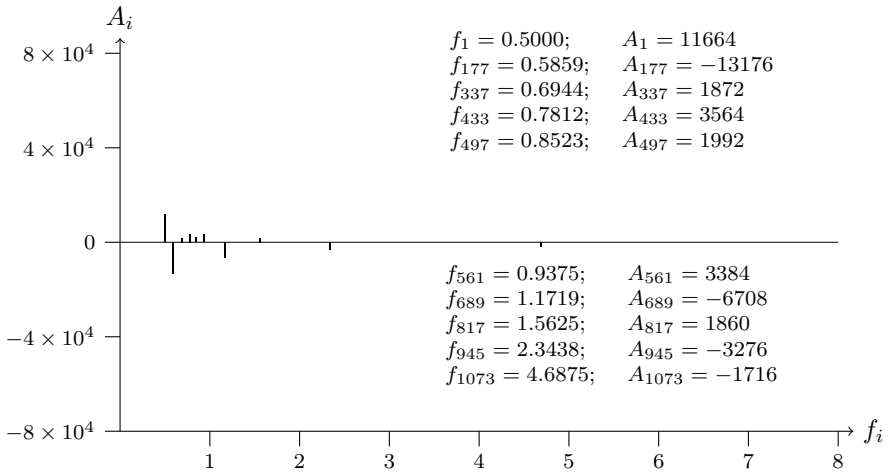

Figure 15a: T2T-CHM13v2.0 chromosome 14; p telomere; analyzed data from 1 to 1200

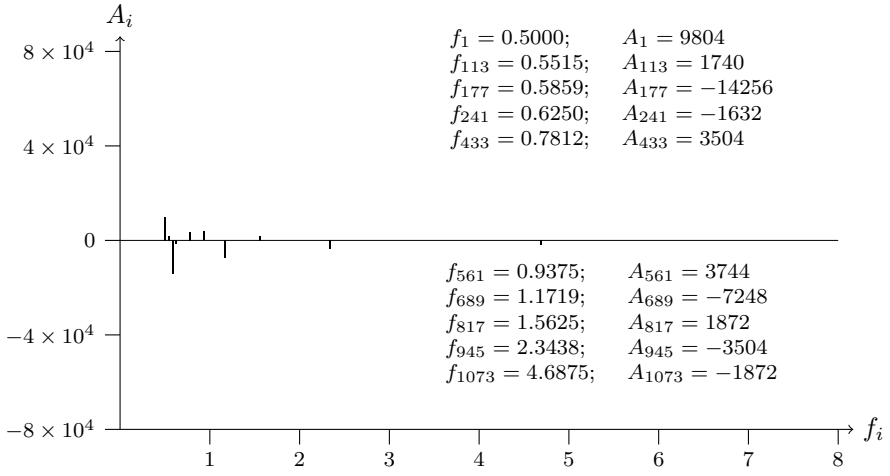

Figure 15b: T2T-CHM13v2.0 chromosome 14; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

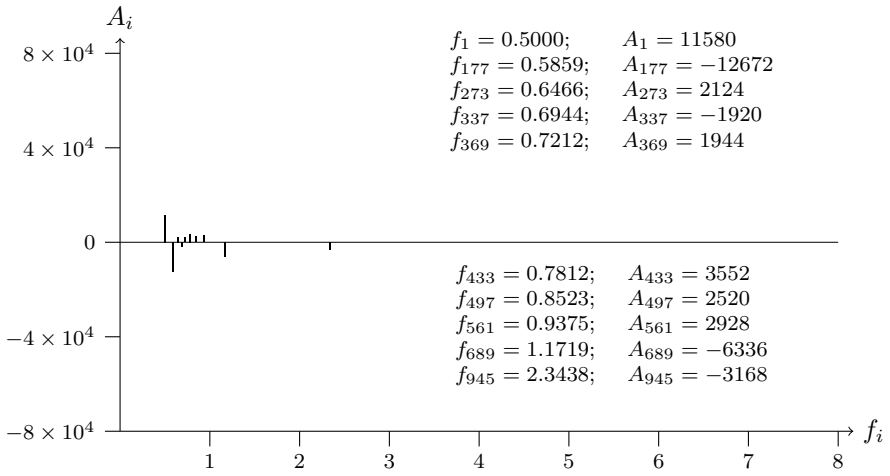

Figure 16a: T2T-CHM13v2.0 chromosome 15; p telomere; analyzed data from 1 to 1200

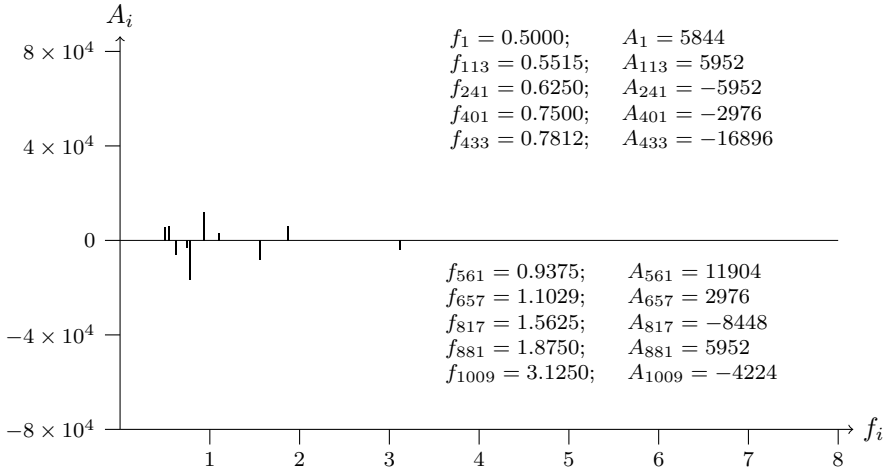

Figure 16b: T2T-CHM13v2.0 chromosome 15; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

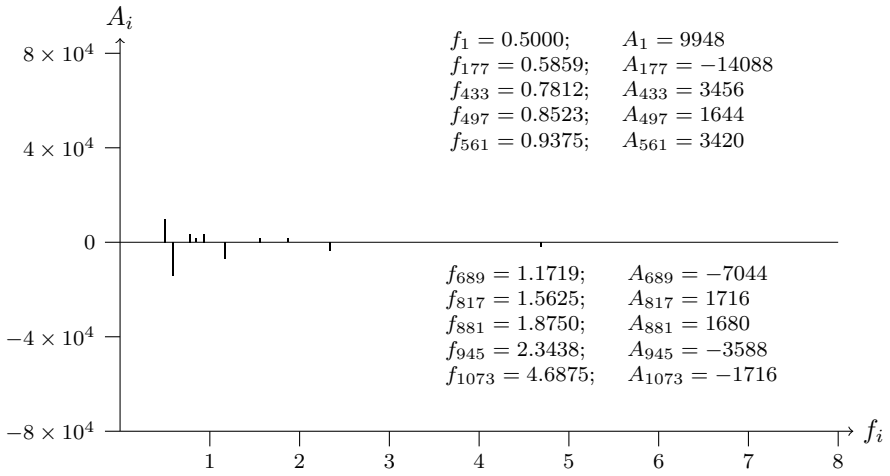

Figure 17a: T2T-CHM13v2.0 chromosome 16; p telomere; analyzed data from 1 to 1200

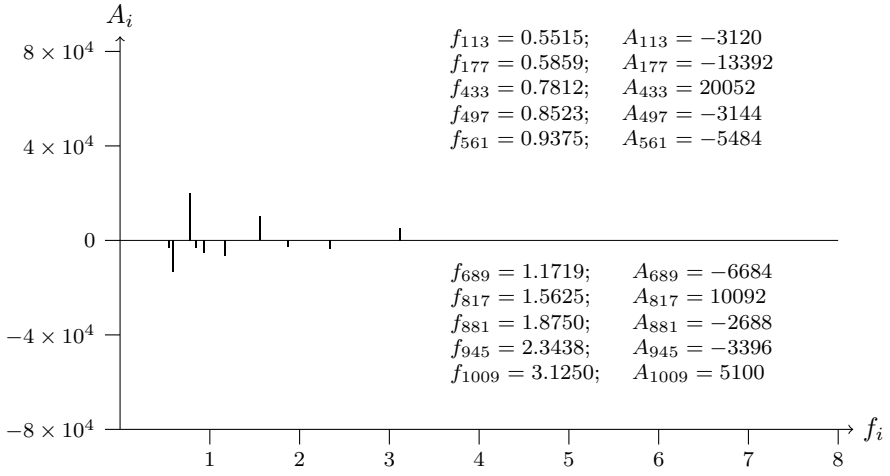

Figure 17b: T2T-CHM13v2.0 chromosome 16; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

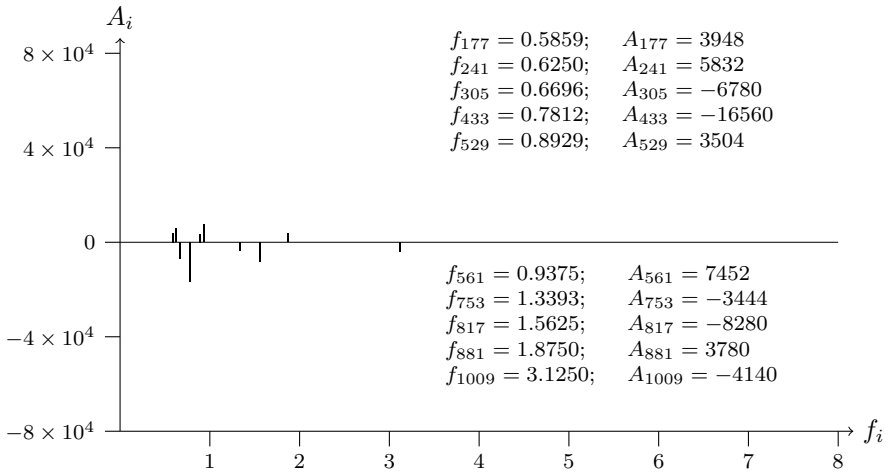

Figure 18a: T2T-CHM13v2.0 chromosome 17; p telomere; analyzed data from 1 to 1200

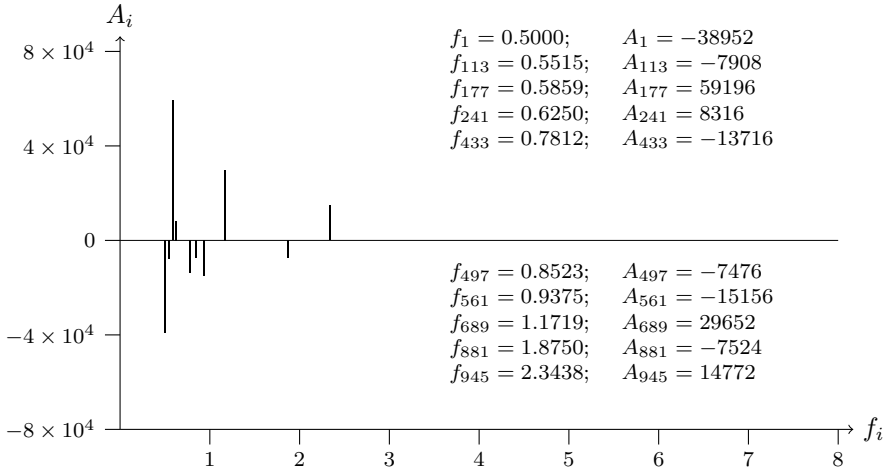

Figure 18b: T2T-CHM13v2.0 chromosome 17; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

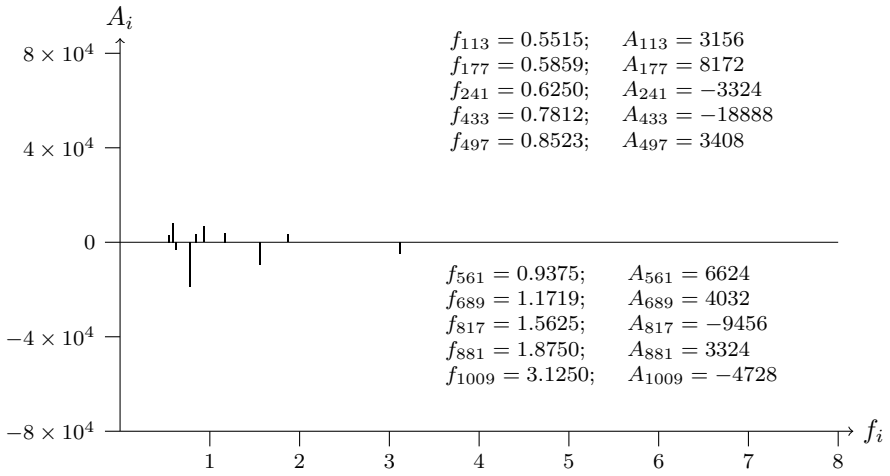

Figure 19a: T2T-CHM13v2.0 chromosome 18; p telomere; analyzed data from 1 to 1200

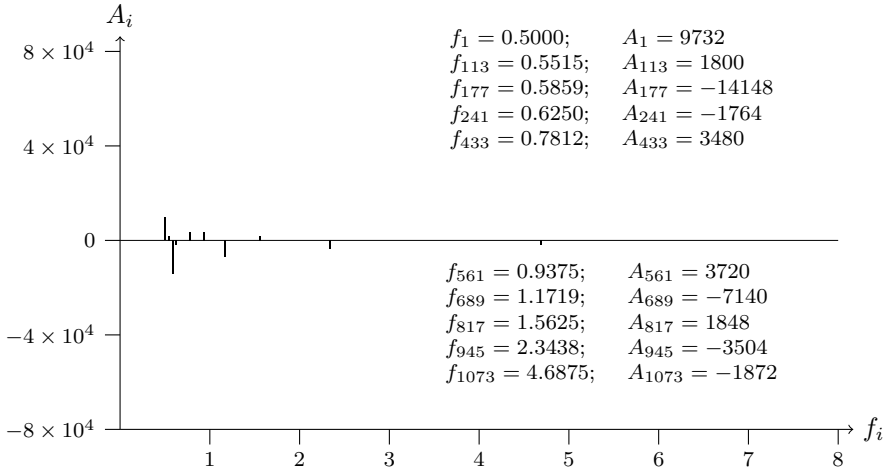

Figure 19b: T2T-CHM13v2.0 chromosome 18; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

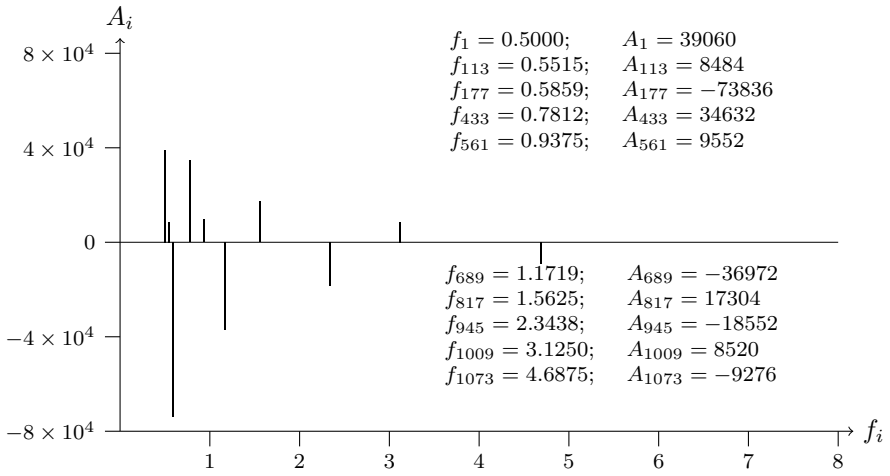

Figure 20a: T2T-CHM13v2.0 chromosome 19; p telomere; analyzed data from 1 to 1200

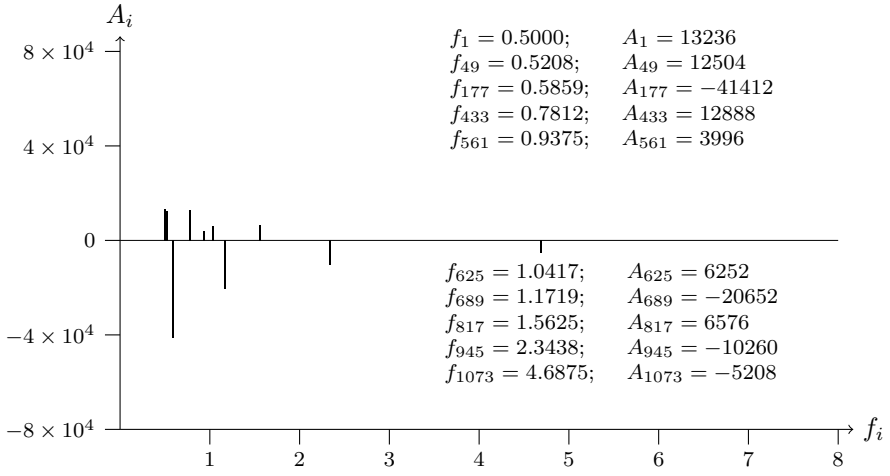

Figure 20b: T2T-CHM13v2.0 chromosome 19; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

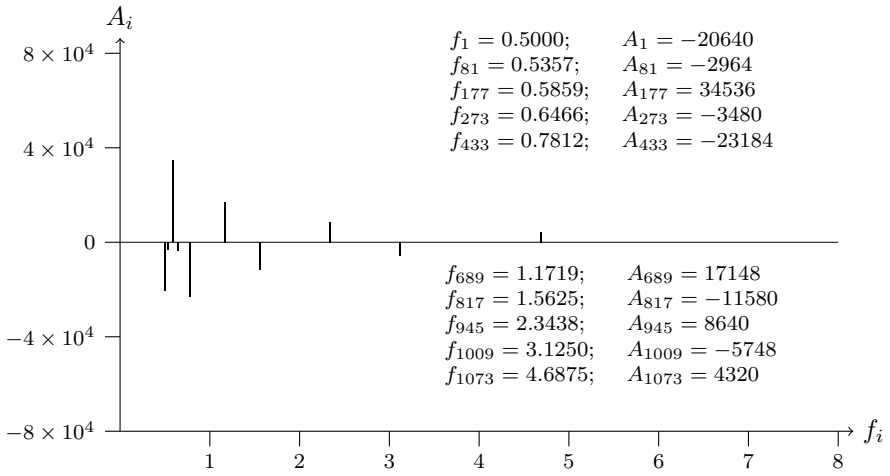

Figure 21a: T2T-CHM13v2.0 chromosome 20; p telomere; analyzed data from 1 to 1200

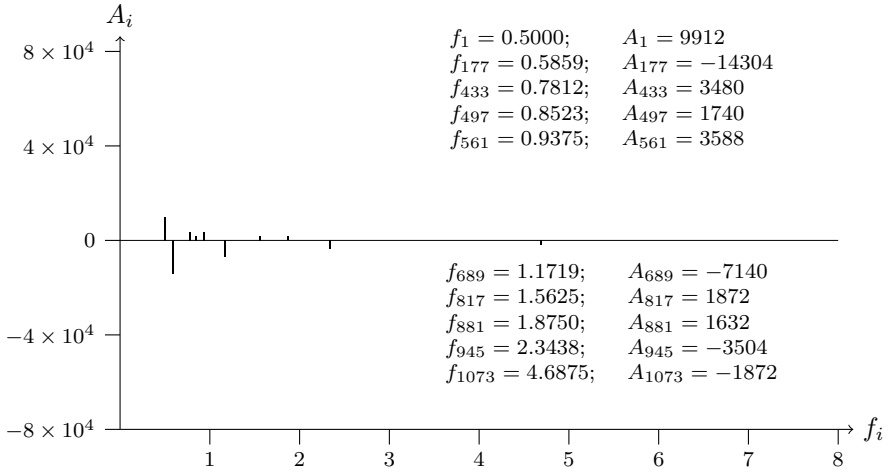

Figure 21b: T2T-CHM13v2.0 chromosome 20; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

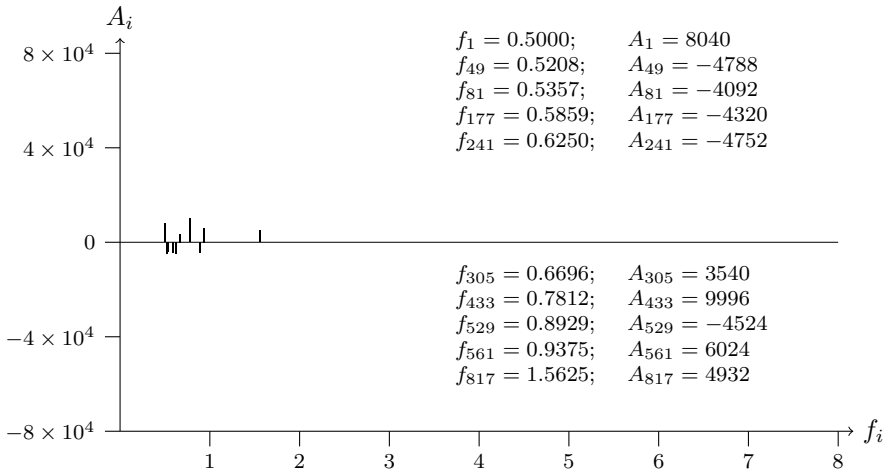

Figure 22a: T2T-CHM13v2.0 chromosome 21; p telomere; analyzed data from 1 to 1200

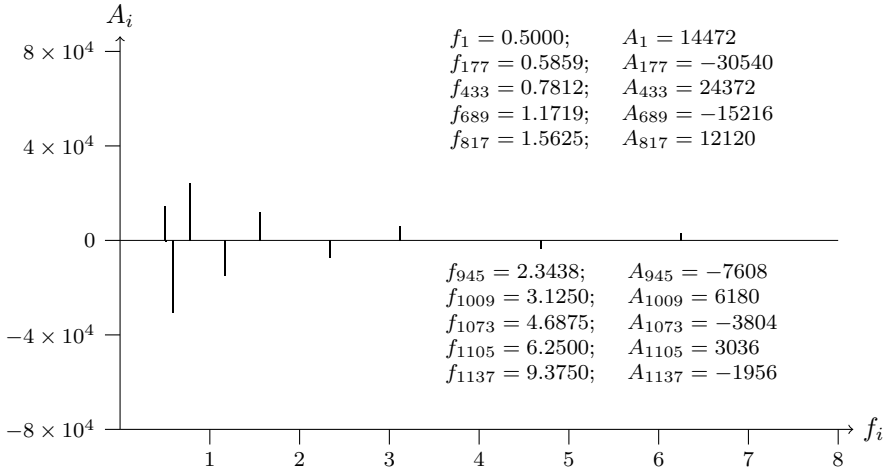

Figure 22b: T2T-CHM13v2.0 chromosome 21; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

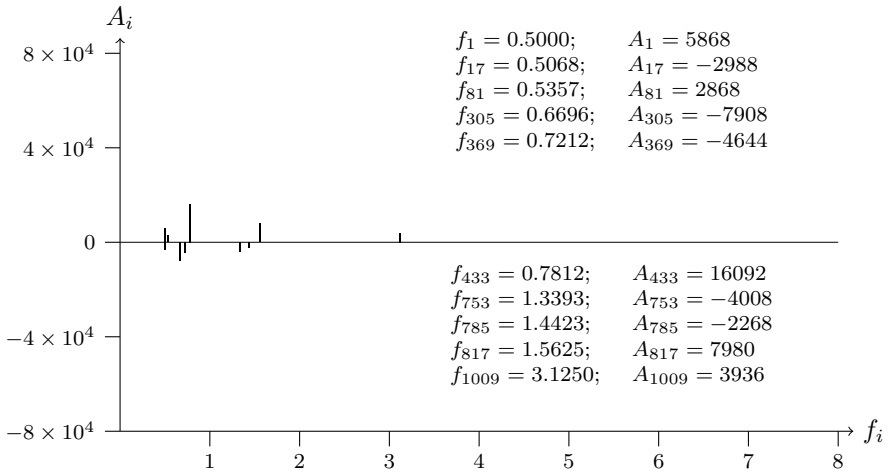

Figure 23a: T2T-CHM13v2.0 chromosome 22; p telomere; analyzed data from 1 to 1200

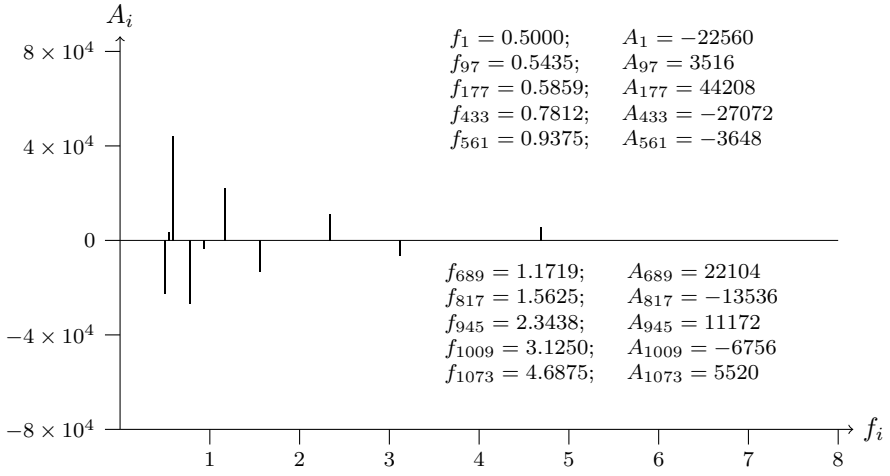

Figure 23b: T2T-CHM13v2.0 chromosome 22; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

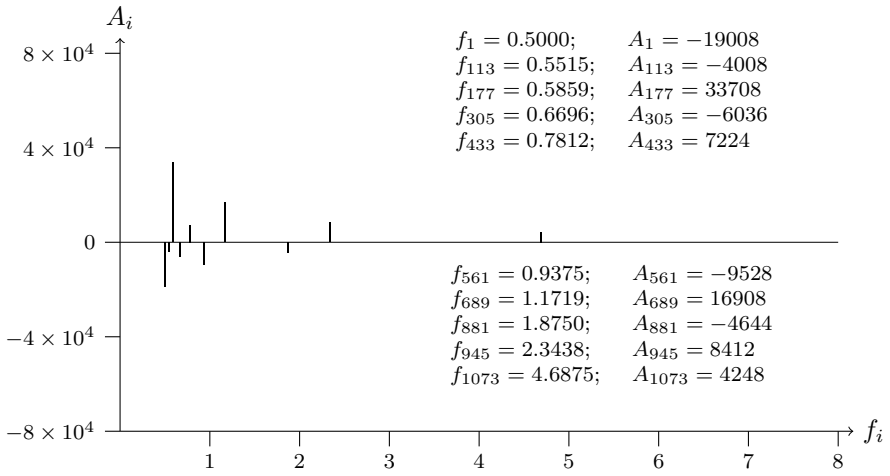

Figure 24a: T2T-CHM13v2.0 chromosome X; p telomere; analyzed data from 1 to 1200

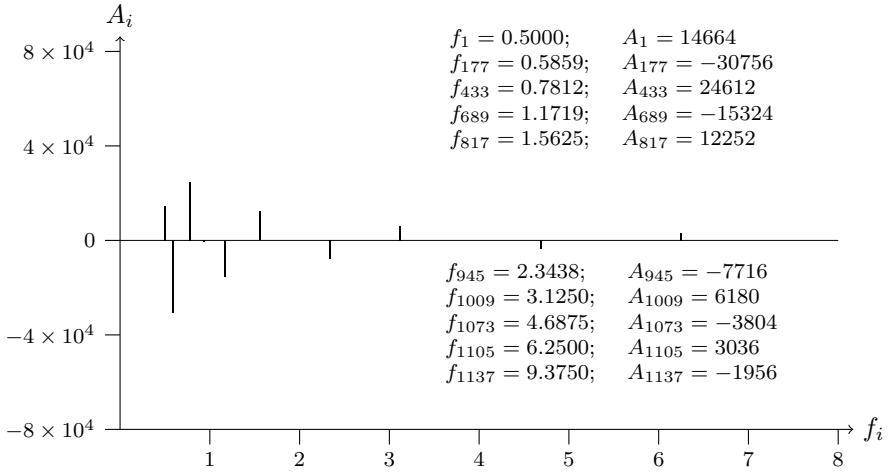

Figure 24b: T2T-CHM13v2.0 chromosome X; p telomere; analyzed data from 1200 to 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

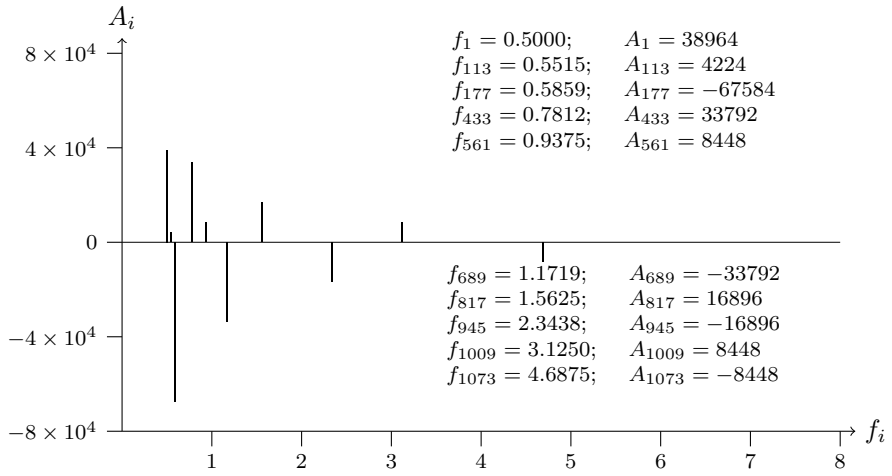

Figure 25a: T2T-CHM13v2.0 chromosome Y; p telomere; analyzed data from 1 to 1200

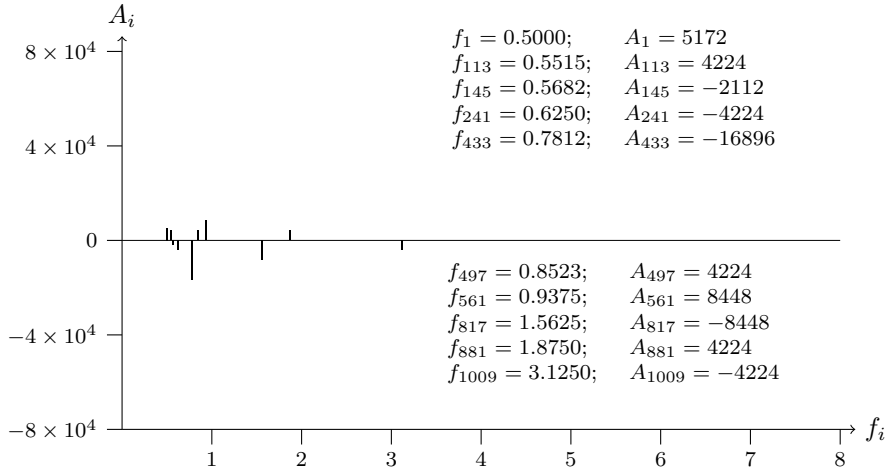

Figure 25b: T2T-CHM13v2.0 chromosome Y; p telomere; analyzed data from 1200 to 1

3.1 SWT analysis of chromosome 1

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

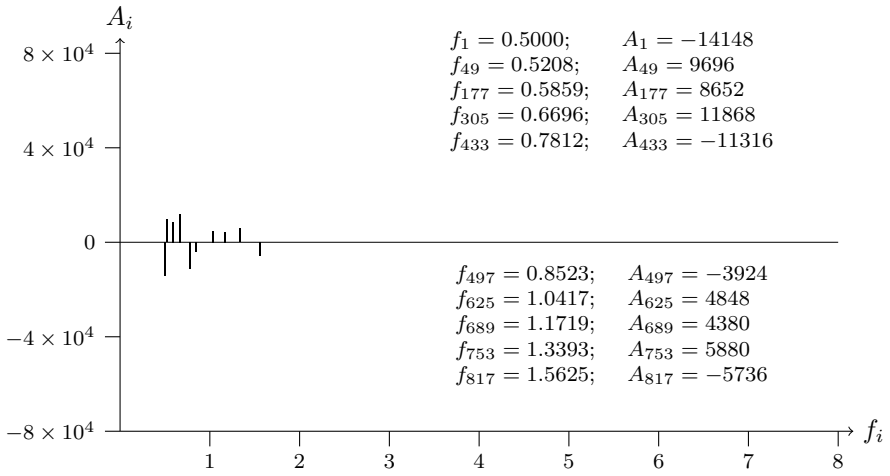

Figure 26a: T2T-CHM13v2.0; chromosome 1; p centromere; analyzed data from 0 to 1199

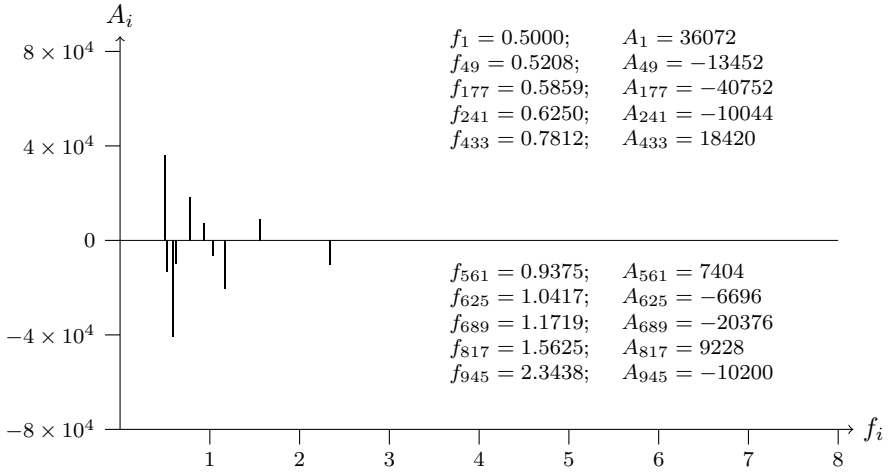

Figure 26b: T2T-CHM13v2.0; chromosome 1; p centromere; analyzed data from 0 to 1199

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

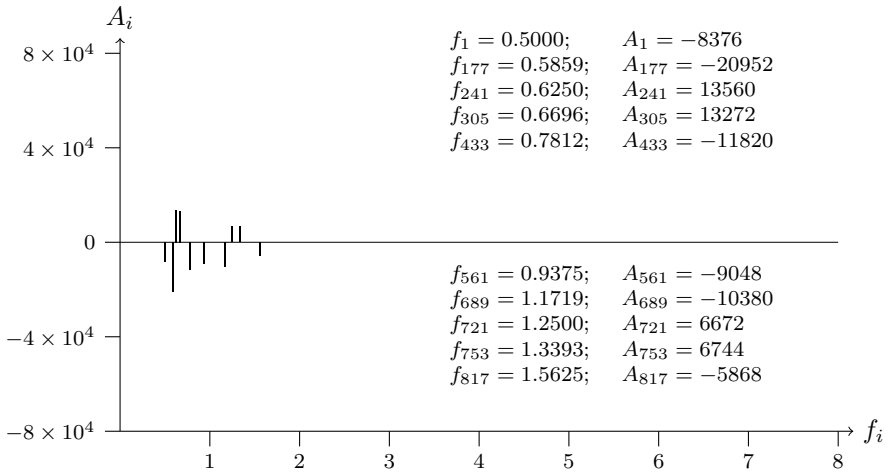

Figure 27a: T2T-CHM13v2.0; chromosome 1; p centromere; analyzed data from 1126109 to 1127308

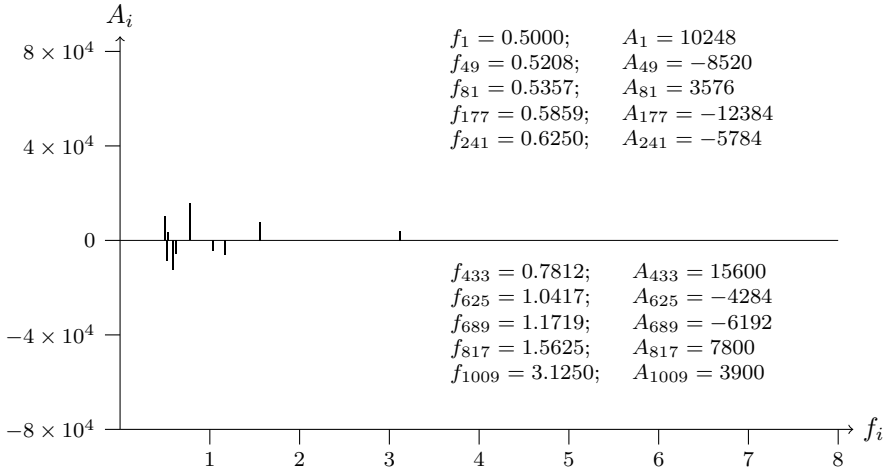

Figure 27b: T2T-CHM13v2.0; chromosome 1; p centromere; analyzed data from 1126109 to 1127308

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

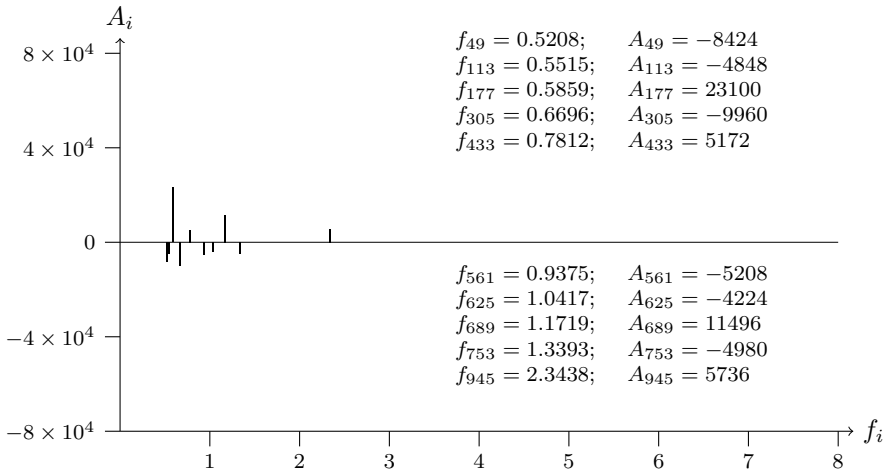

Figure 28a: T2T-CHM13v2.0; chromosome 1; p centromere; analyzed data from 2251019 to 2252219

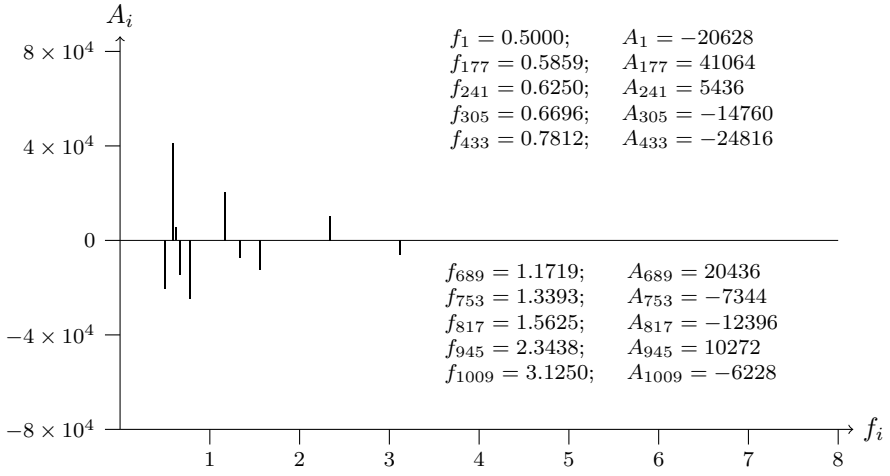

Figure 28b: T2T-CHM13v2.0; chromosome 1; p centromere; analyzed data from 2251019 to 2252219

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

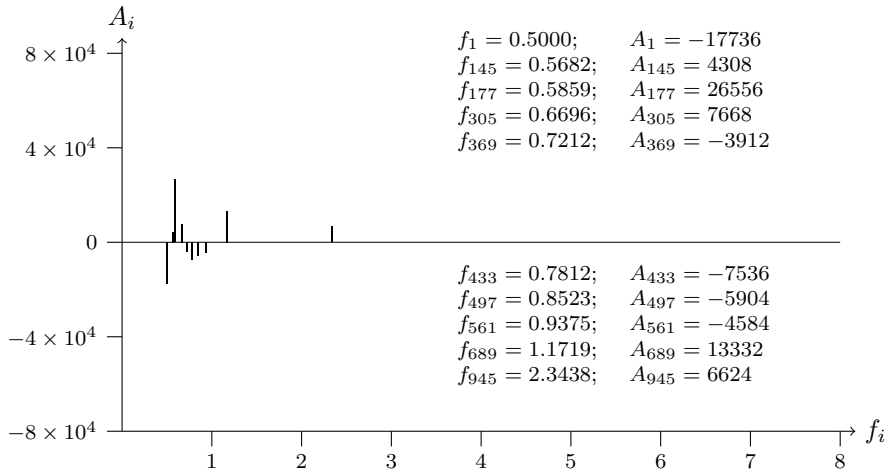

Figure 29a: T2T-CHM13v2.0; chromosome 1; q centromere; analyzed data from 0 to 1199

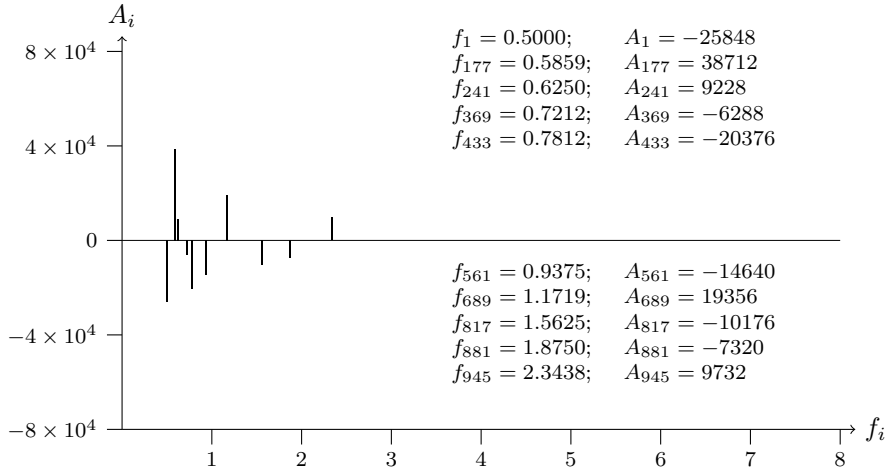

Figure 29b: T2T-CHM13v2.0; chromosome 1; q centromere; analyzed data from 0 to 1199

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

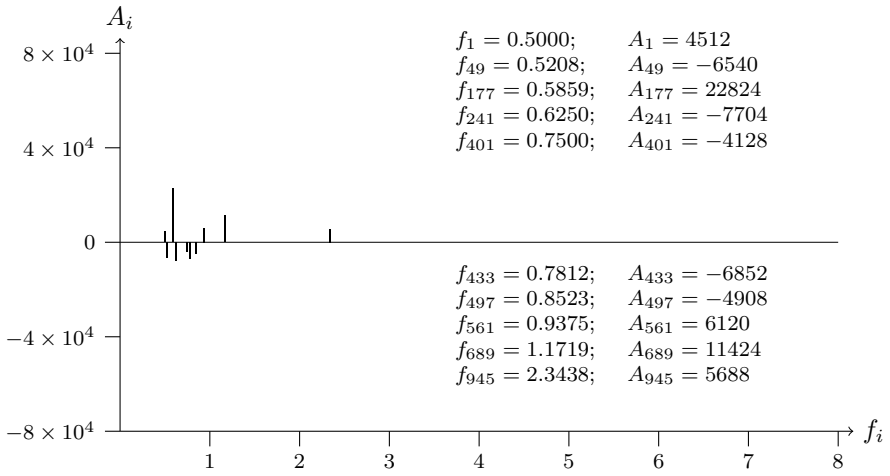

Figure 30a: T2T-CHM13v2.0; chromosome 1; q centromere; analyzed data from 1126110 to 1127309

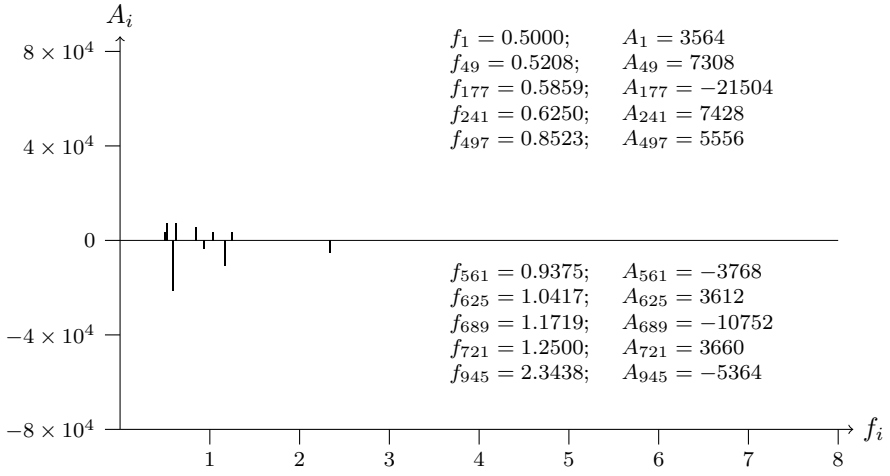

Figure 30b: T2T-CHM13v2.0; chromosome 1; q centromere; analyzed data from 1126110 to 1127309

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

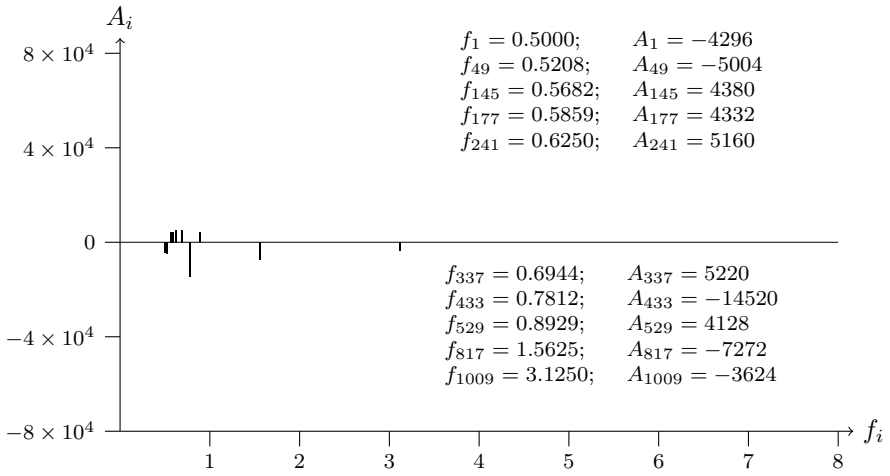

Figure 31a: T2T-CHM13v2.0; chromosome 1; q centromere; analyzed data from 2251020 to 2252220

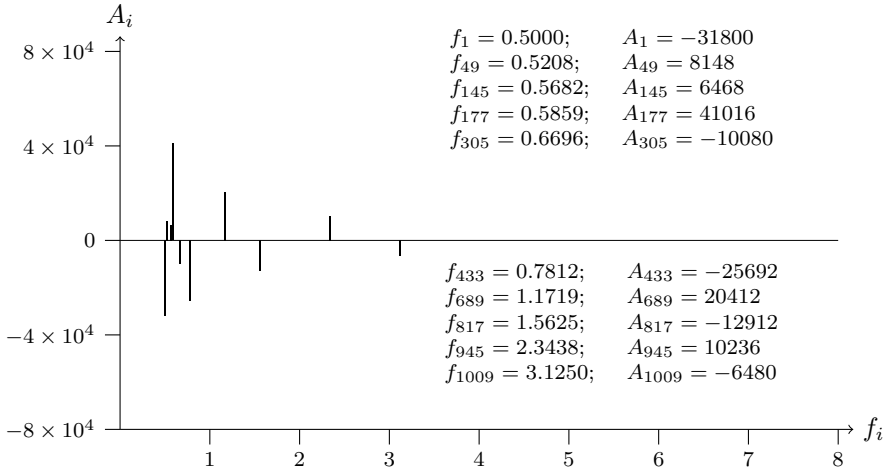

Figure 31b: T2T-CHM13v2.0; chromosome 1; q centromere; analyzed data from 2251020 to 2252220

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

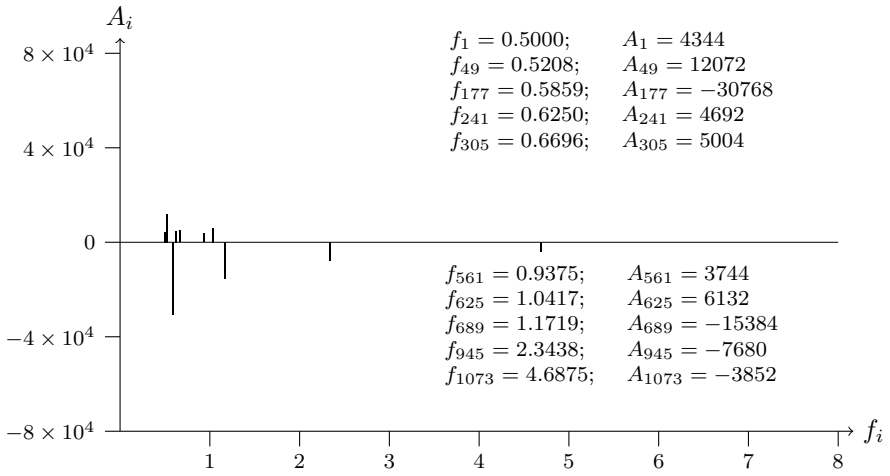

Figure 32a: T2T-CHM13v2.0; chromosome 1; p telomere; analyzed data from 0 to 1199

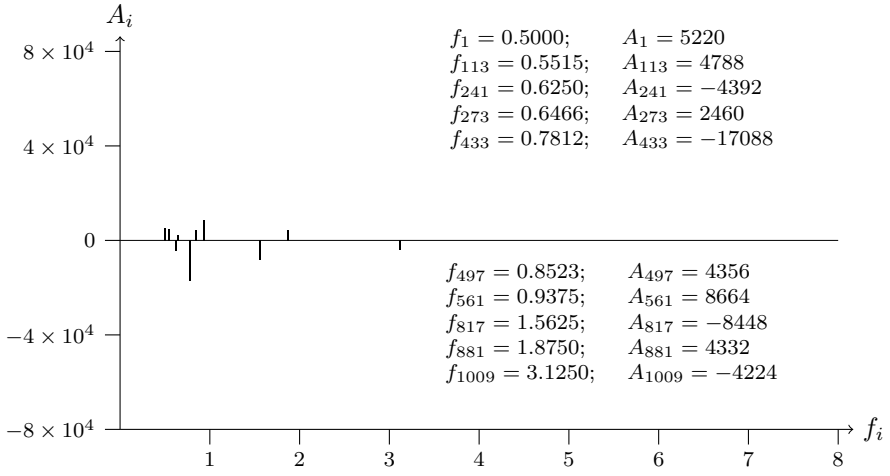

Figure 32b: T2T-CHM13v2.0; chromosome 1; p telomere; analyzed data from 0 to 1199

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

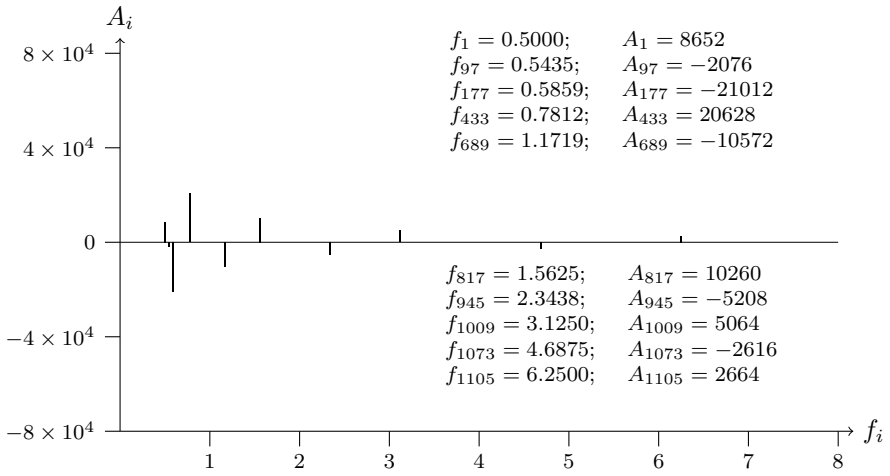

Figure 33a: T2T-CHM13v2.0; chromosome 1; p telomere; analyzed data from 1500 to 2699

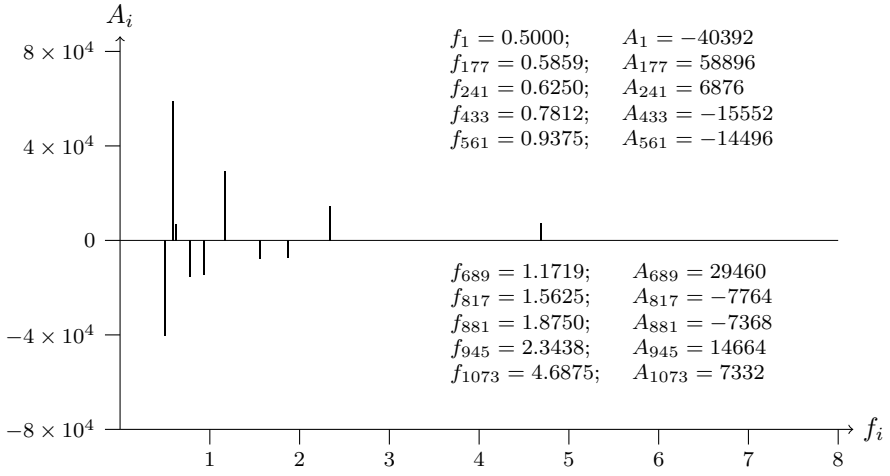

Figure 33b: T2T-CHM13v2.0; chromosome 1; p telomere; analyzed data from 1500 to 2699

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

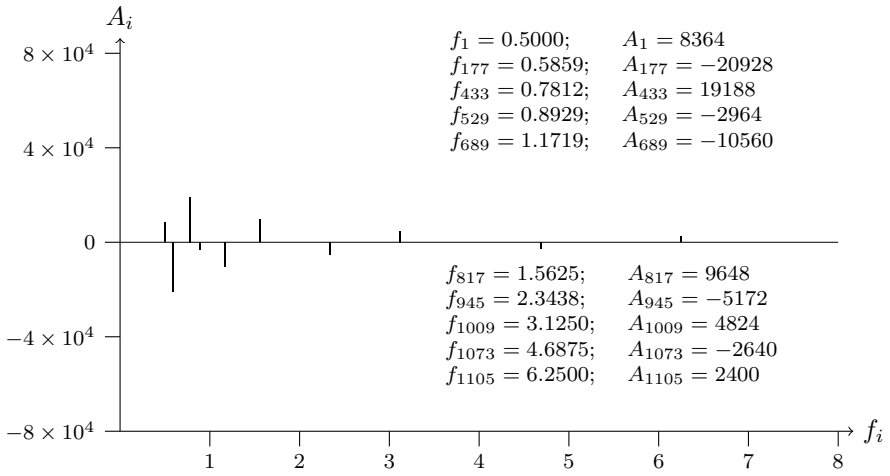

Figure 34a: T2T-CHM13v2.0; chromosome 1; p telomere; analyzed data from 1800 to 3000

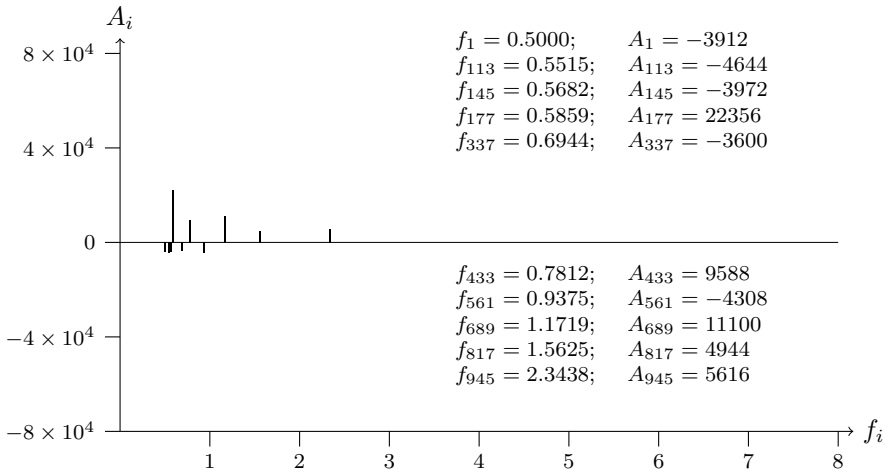

Figure 34b: T2T-CHM13v2.0; chromosome 1; p telomere; analyzed data from 1800 to 3000

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

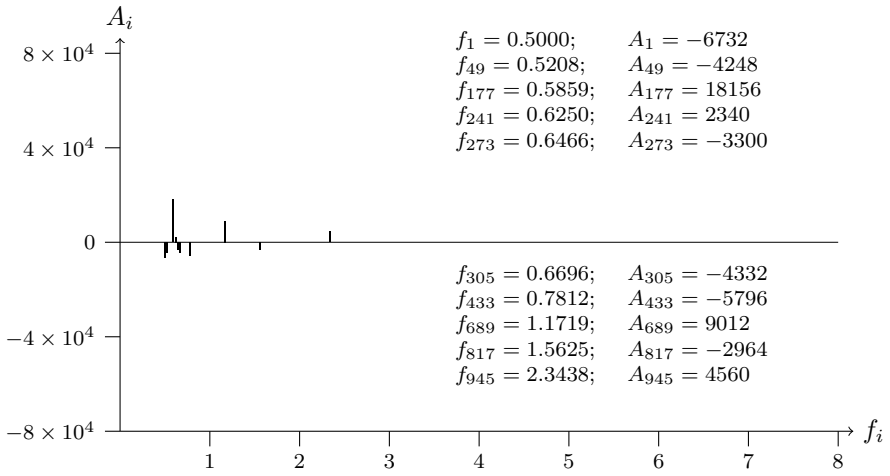

Figure 35a: T2T-CHM13v2.0; chromosome 1; q telomere; analyzed data from 0 to 1199

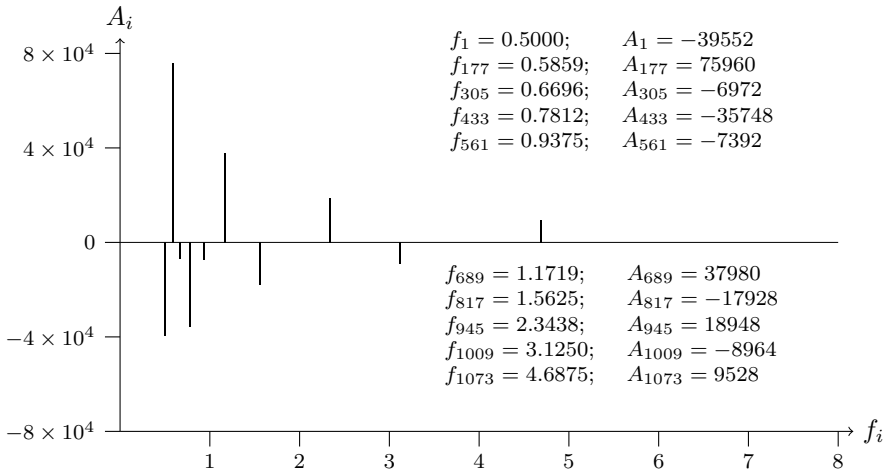

Figure 35b: T2T-CHM13v2.0; chromosome 1; q telomere; analyzed data from 0 to 1199

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

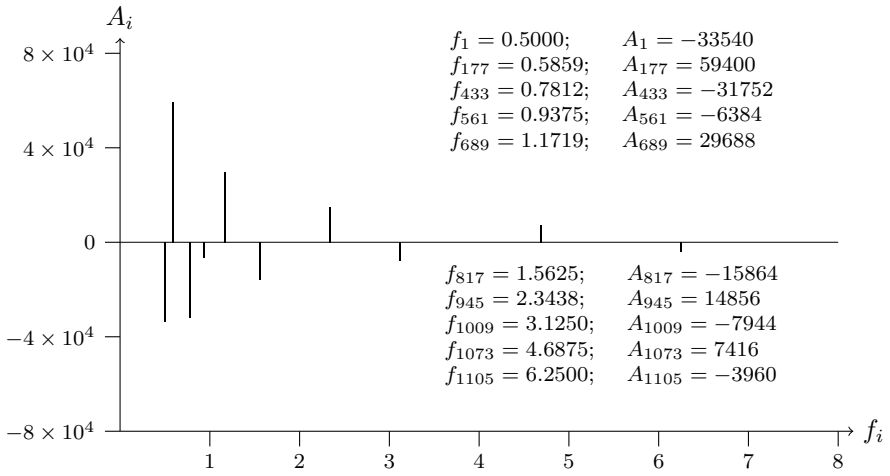

Figure 36a: T2T-CHM13v2.0; chromosome 1; q telomere; analyzed data from 1664 to 2863

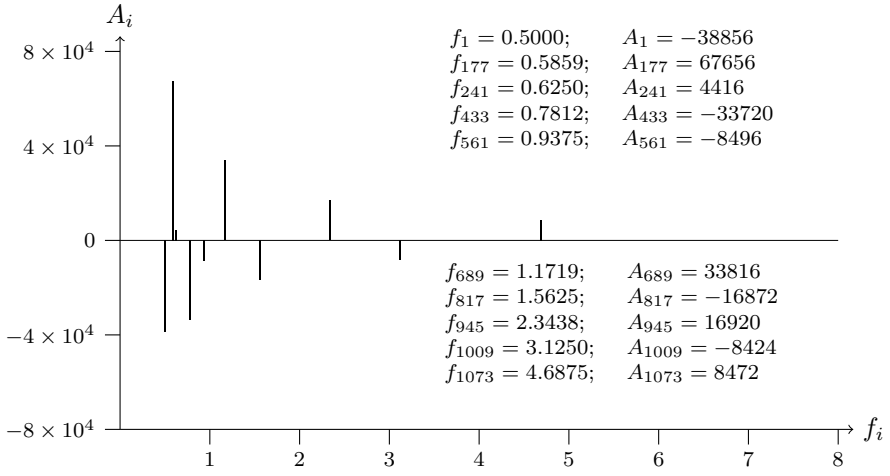

Figure 36b: T2T-CHM13v2.0; chromosome 1; q telomere; analyzed data from 1664 to 2863

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

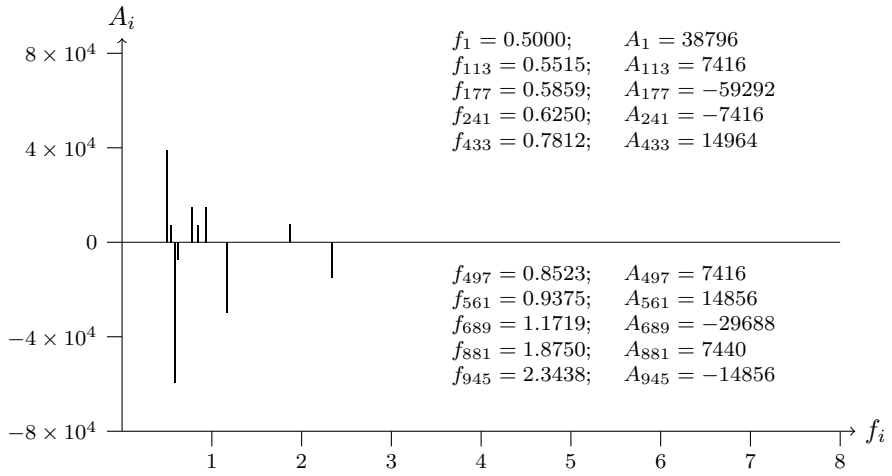

Figure 37a: T2T-CHM13v2.0; chromosome 1; q telomere; analyzed data from 2128 to 3328

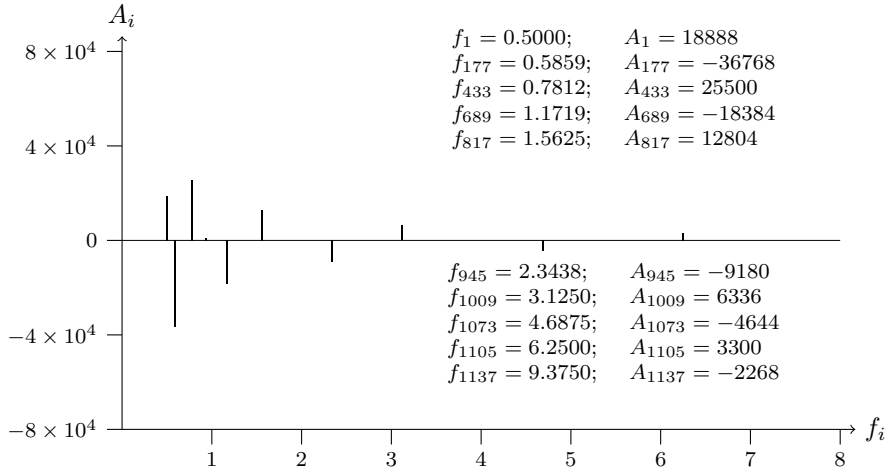

Figure 37b: T2T-CHM13v2.0; chromosome 1; q telomere; analyzed data from 2128 to 3328

4 Biological analysis SWT for genes

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

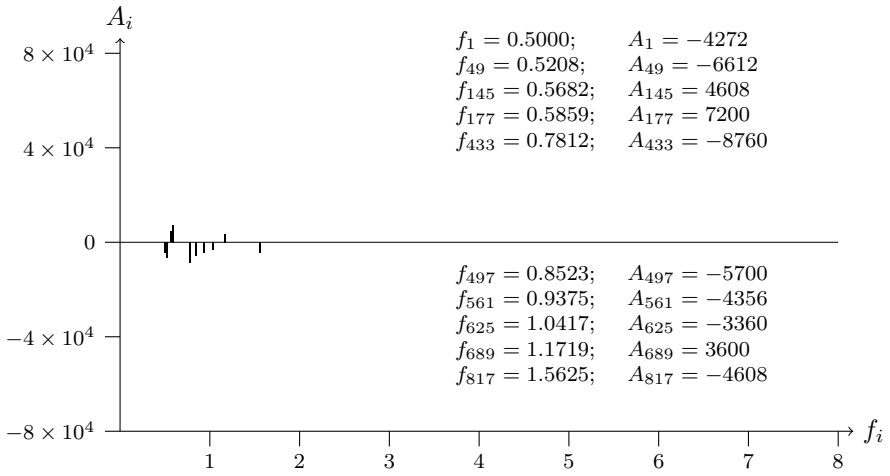

Figure 38a: T2T-CHM13v2.0; chromosome 1; RUSC1 (oncogene) exon-NM_001105203.2-2 analyzed data from 154460228 to 154461428

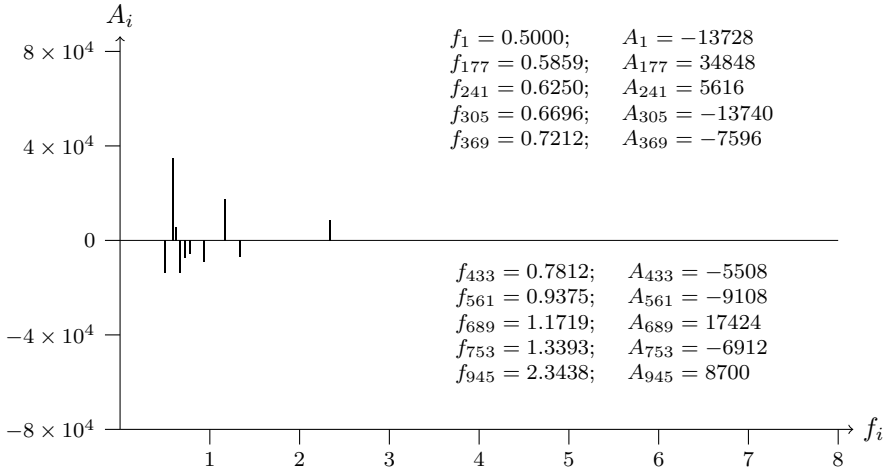

Figure 38b: T2T-CHM13v2.0; chromosome 1; RUSC1 (oncogene) exon-NM_001105203.2-2 analyzed data from 154461428 to 154460228

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

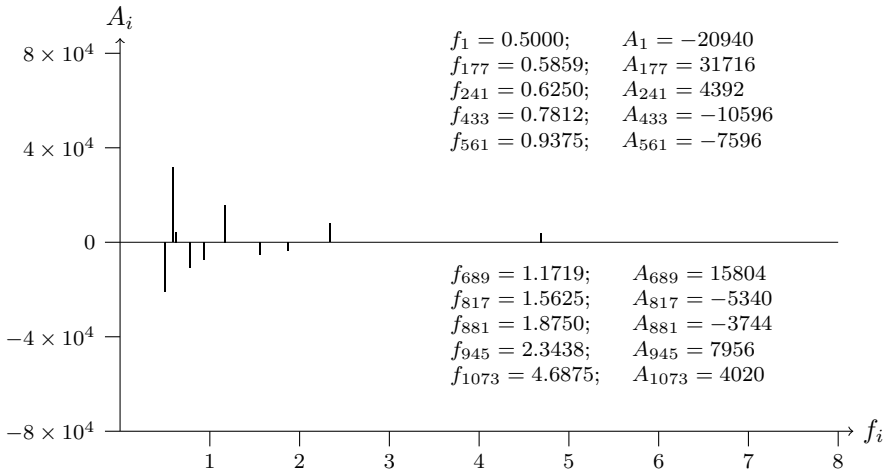

Figure 39a: T2T-CHM13v2.0; chromosome 1; RUSC1 (oncogene) intron-NM_001105203.2-2 analyzed data from 154461670 to 154462870

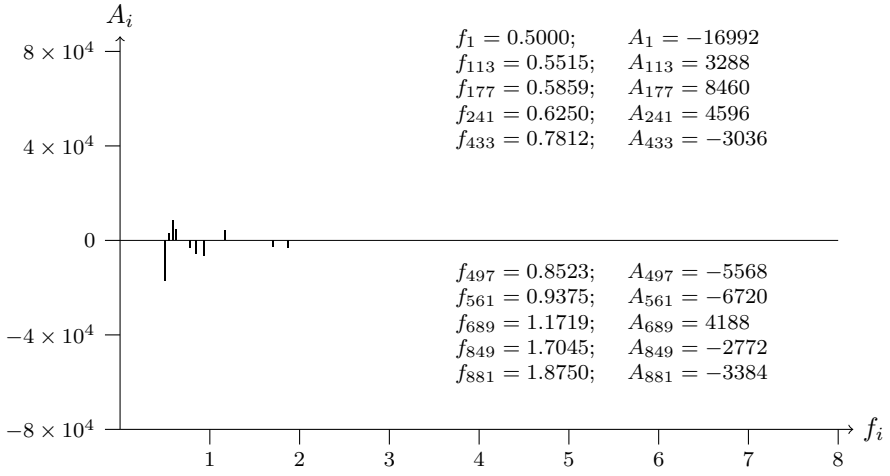

Figure 39b: T2T-CHM13v2.0; chromosome 1; RUSC1 (oncogene) intron-NM_001105203.2-2 analyzed data from 154462870 to 154461670

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

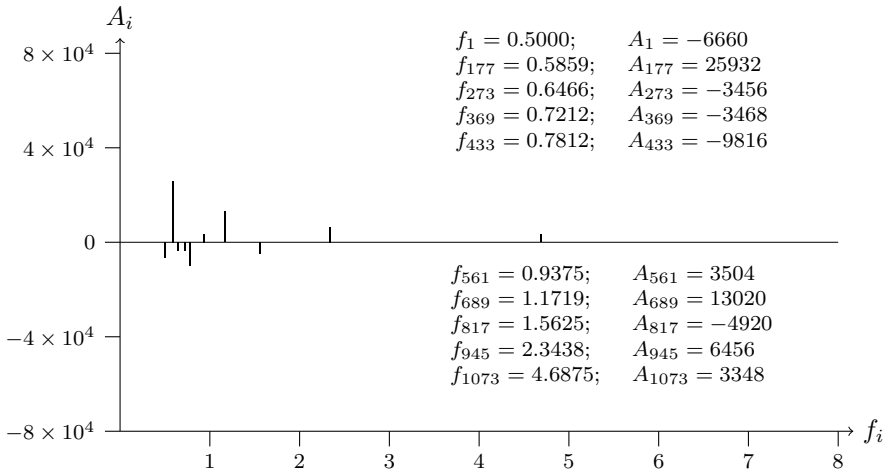

Figure 40a: T2T-CHM13v2.0; chromosome 1; RUSC1 (oncogene) intron-NM_001105203.2-9 analyzed data from 154466815 to 154468015

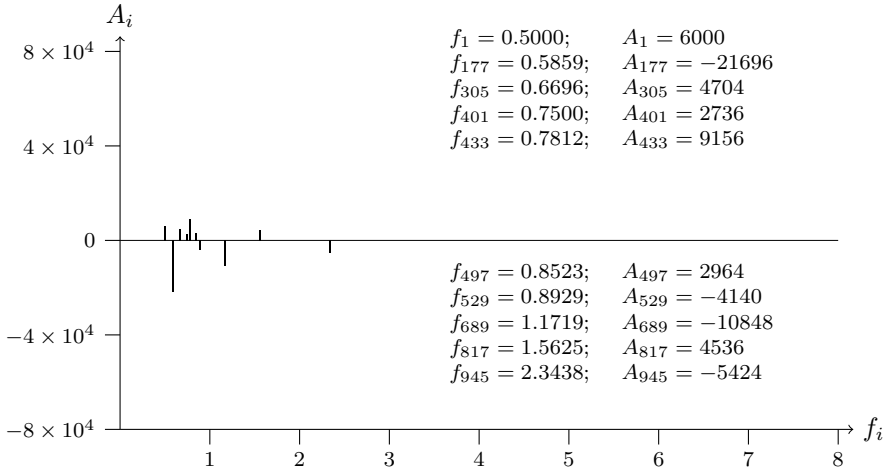

Figure 40b: T2T-CHM13v2.0; chromosome 1; RUSC1 (oncogene) intron-NM_001105203.2-9 analyzed data from 154468015 to 154466815

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

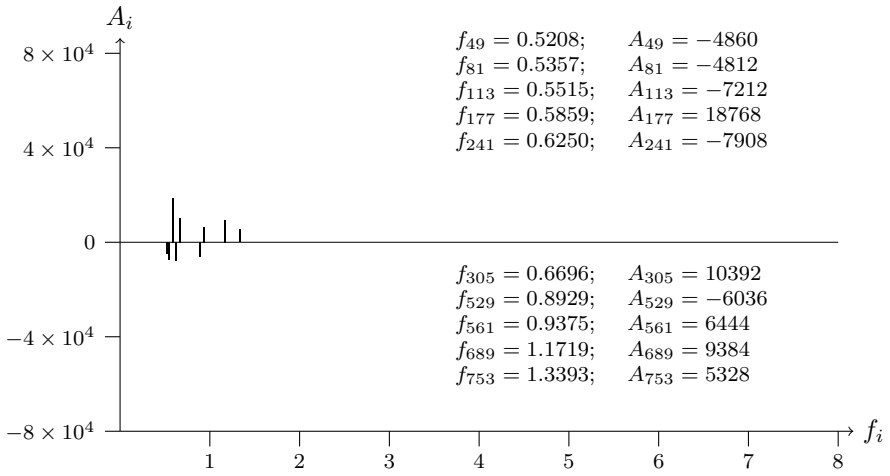

Figure 41a: T2T-CHM13v2.0; chromosome 1; GPR161 exon-NM_001349635.1-5 analyzed data from 167431007 to 167432207

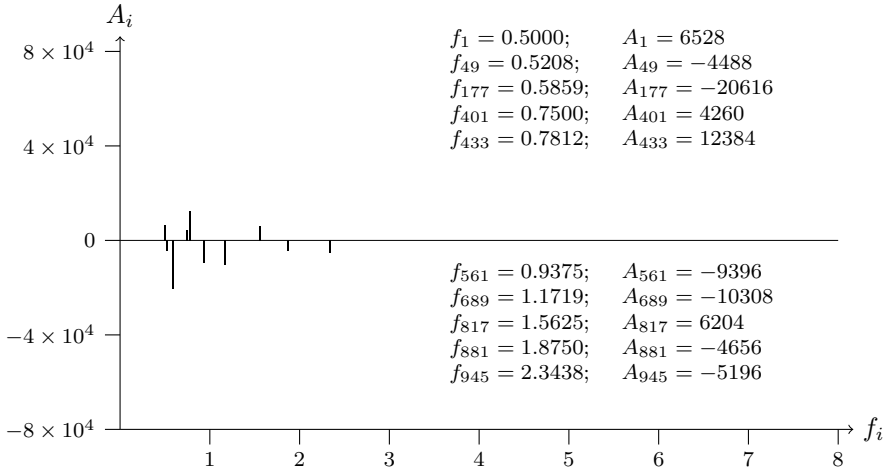

Figure 41b: T2T-CHM13v2.0; chromosome 1; GPR161 exon-NM_001349635.1-5 analyzed data from 167432207 to 167431007

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

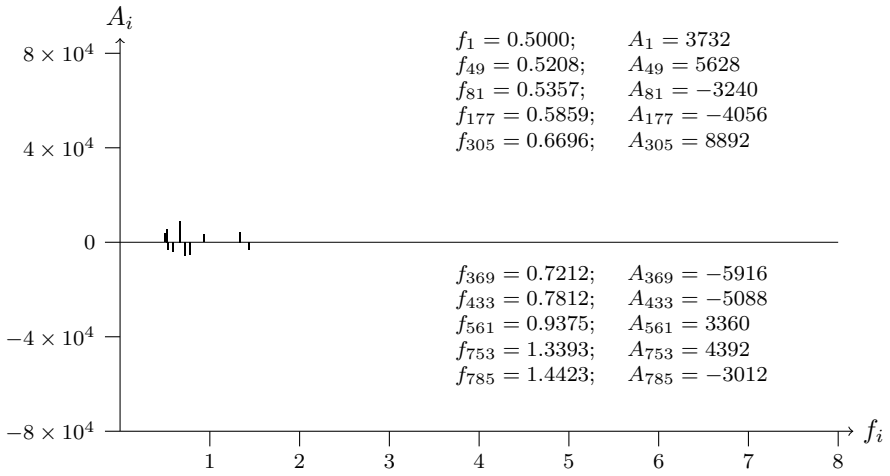

Figure 42a: T2T-CHM13v2.0; chromosome 17; BRCA1 (tumor supressor) exon-NM_001408458.1-22 analyzed data from 43902857 to 43904057

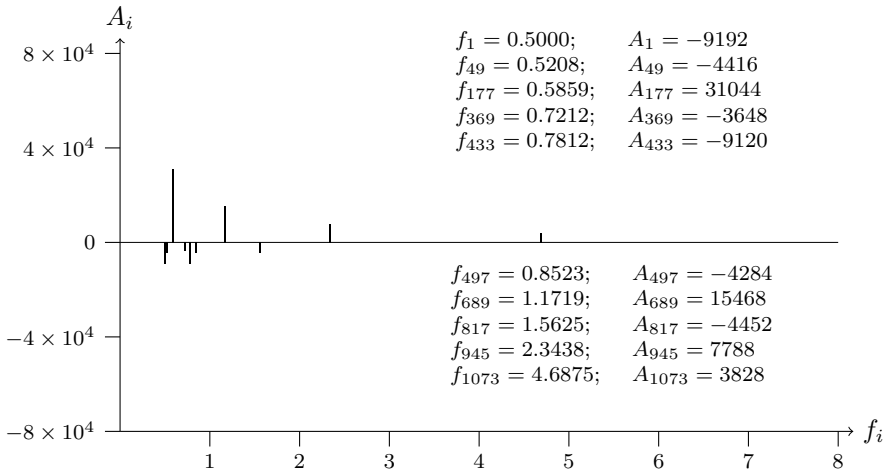

Figure 42b: T2T-CHM13v2.0; chromosome 17; BRCA1 (tumor supressor) exon-NM_001408458.1-22 analyzed data from 43904057 to 43902857

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

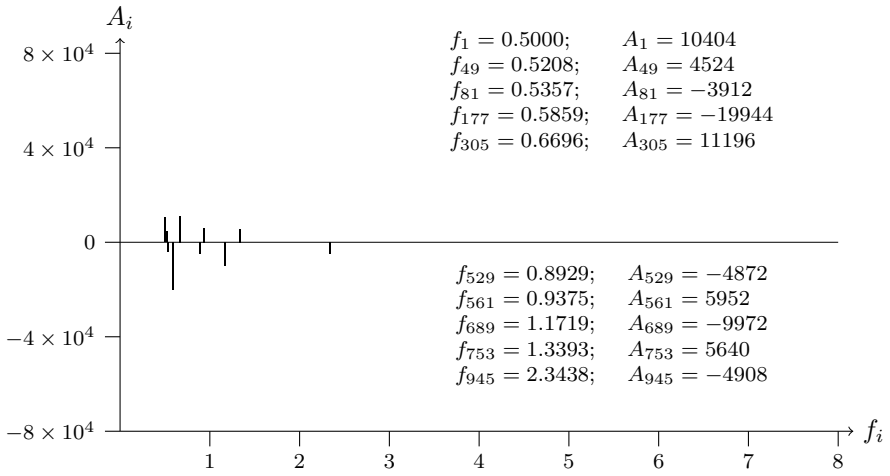

Figure 43a: T2T-CHM13v2.0; chromosome 17; BRCA1 (tumor supressor) intron-NM_001408458.1-22 analyzed data from 43904362 to 43905562

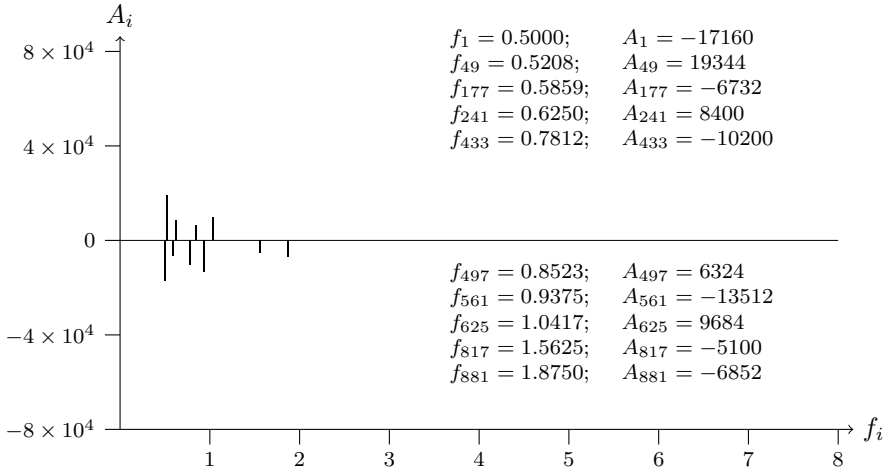

Figure 43b: T2T-CHM13v2.0; chromosome 17; BRCA1 (tumor supressor) intron-NM_001408458.1-22 analyzed data from 43905562 to 43904362

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

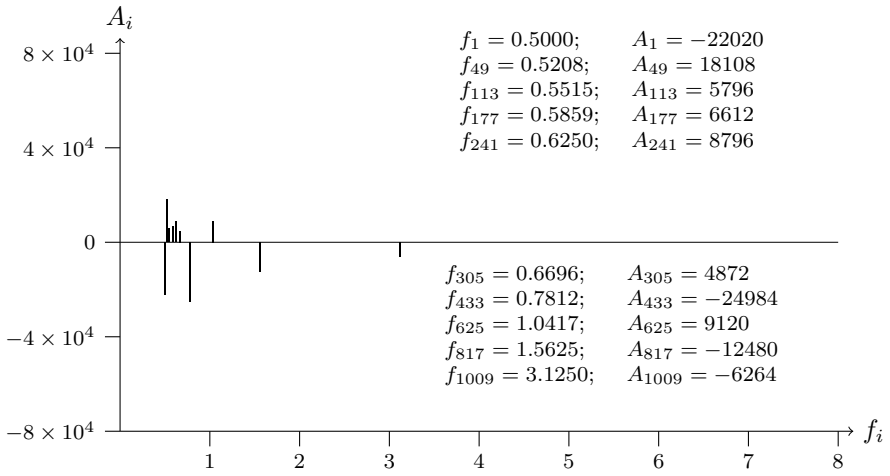

Figure 44a: T2T-CHM13v2.0; chromosome 13; BRCA2 (tumor supressor) exon-NM_000059.4-11 analyzed data from 31553507 to 31554707

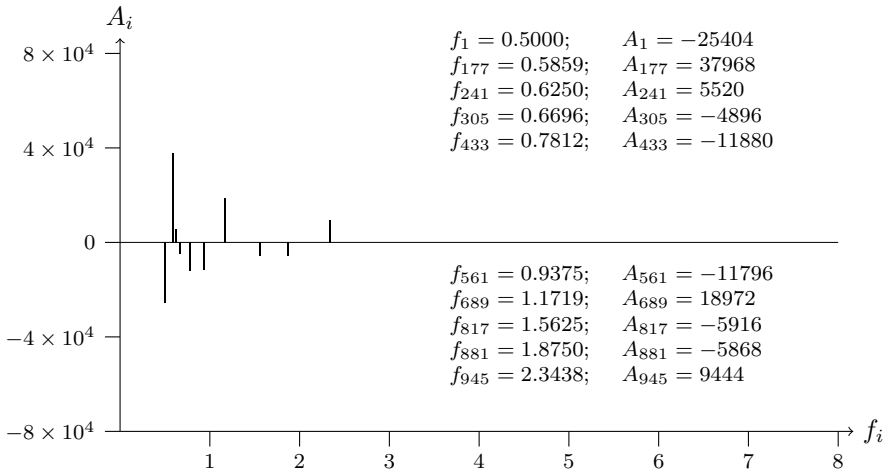

Figure 44b: T2T-CHM13v2.0; chromosome 13; BRCA2 (tumor supressor) exon-NM_000059.4-11 analyzed data from 31554707 to 31553507

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

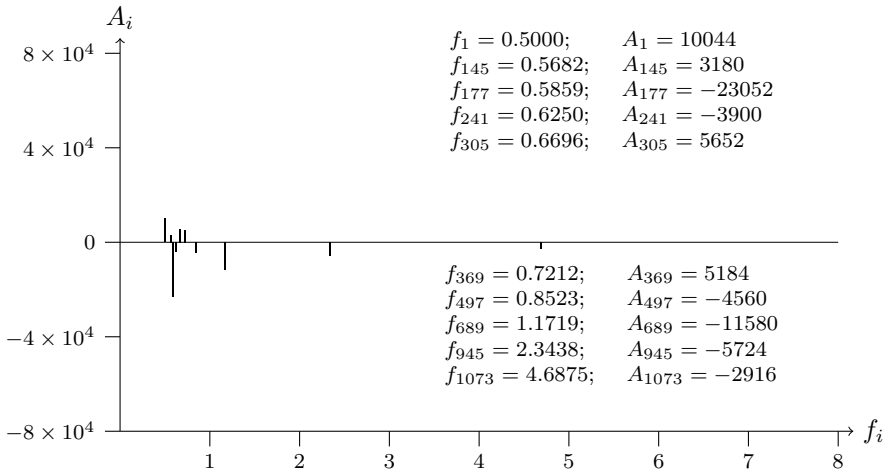

Figure 45a: T2T-CHM13v2.0; chromosome 13; BRCA2 (tumor supressor) exon-NM_000059.4-27 analyzed data from 31615404 to 31616604

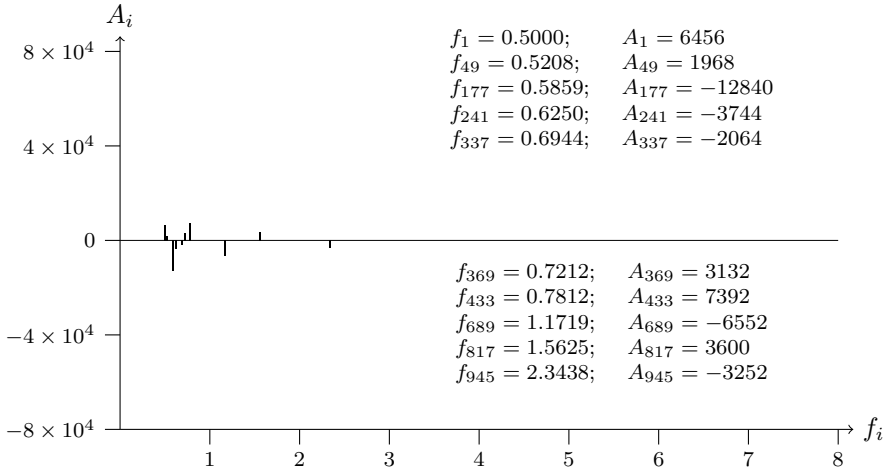

Figure 45b: T2T-CHM13v2.0; chromosome 13; BRCA2 (tumor supressor) exon-NM_000059.4-27 analyzed data from 31616604 to 31615404

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

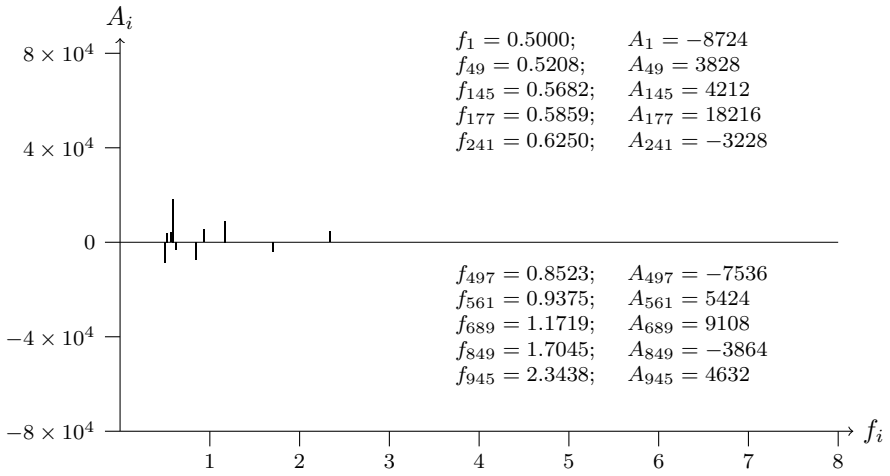

Figure 46a: T2T-CHM13v2.0; chromosome 7; EGFR (oncogene) exon-NM_005228.5-28 analyzed data from 55365684 to 55366884

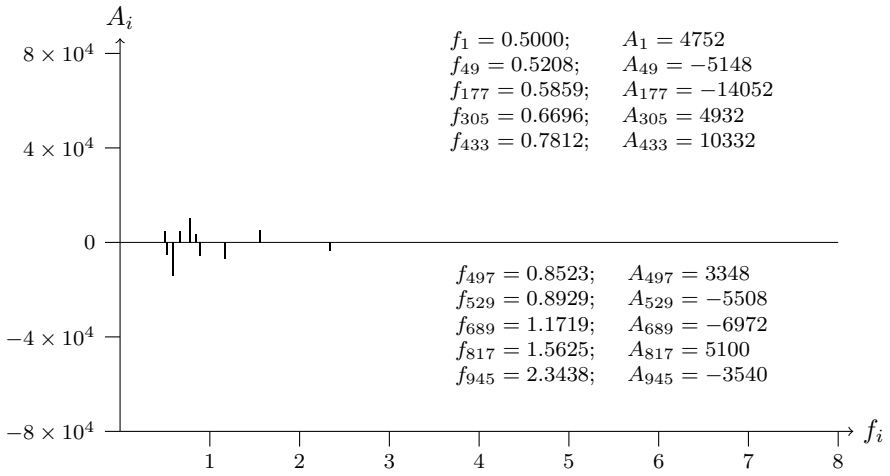

Figure 46b: T2T-CHM13v2.0; chromosome 7; EGFR (oncogene) exon-NM_005228.5-28 analyzed data from 55366884 to 55365684

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

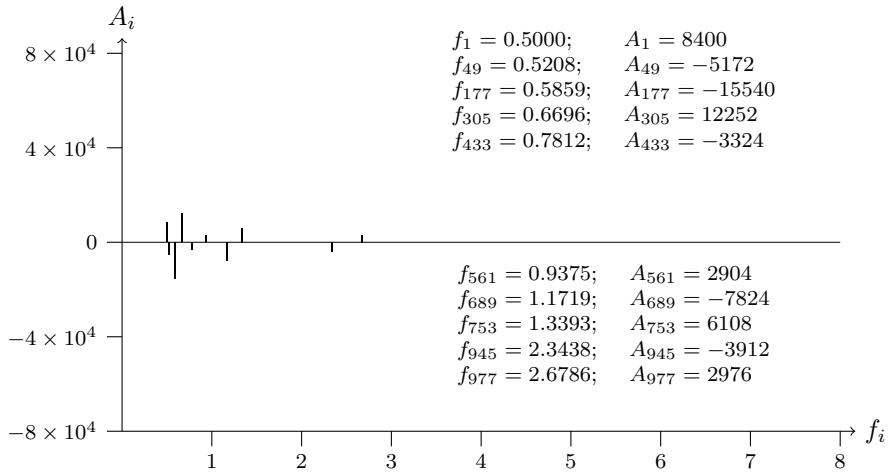

Figure 47a: T2T-CHM13v2.0; chromosome 1; ELK4 (oncogene) exon-NM_001973.4-5 analyzed data from 204872578 to 204873778

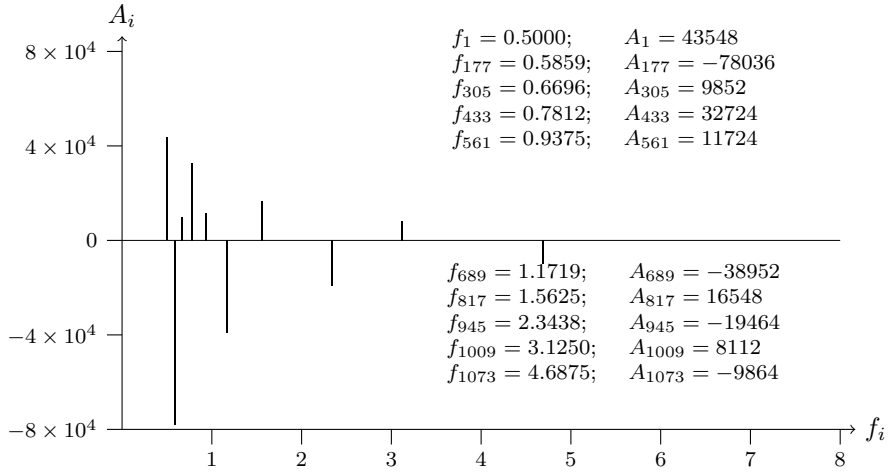

Figure 47b: T2T-CHM13v2.0; chromosome 1; ELK4 (oncogene) exon-NM_001973.4-5 analyzed data from 204873778 to 204872578

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

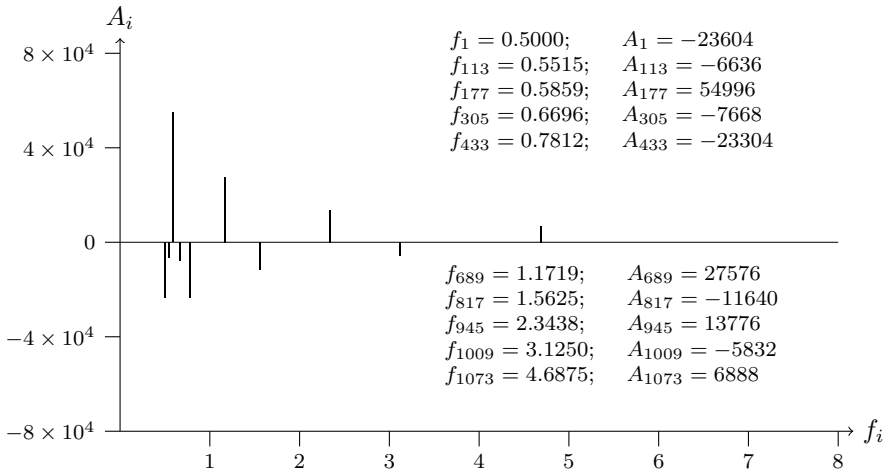

Figure 48a: T2T-CHM13v2.0; chromosome 2; REL (oncogene) exon-NM_002908.4-11 analyzed data from 60927604 to 60928804

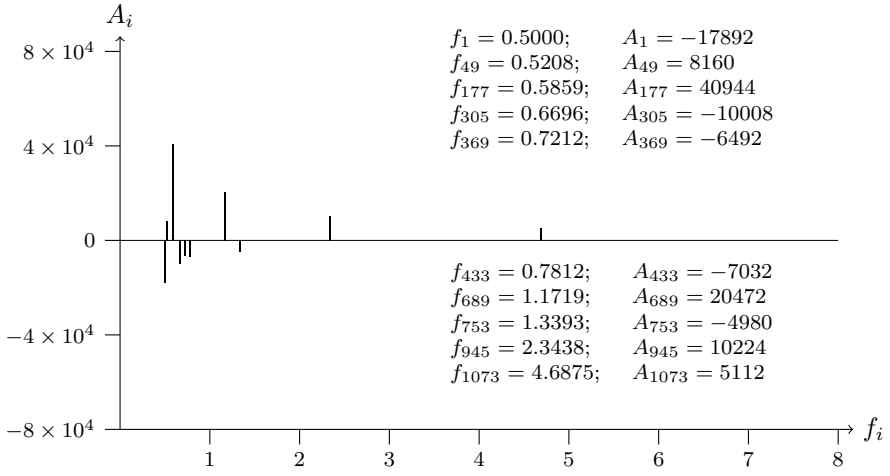

Figure 48b: T2T-CHM13v2.0; chromosome 2; REL (oncogene) exon-NM_002908.4-11 analyzed data from 60928804 to 60927604

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

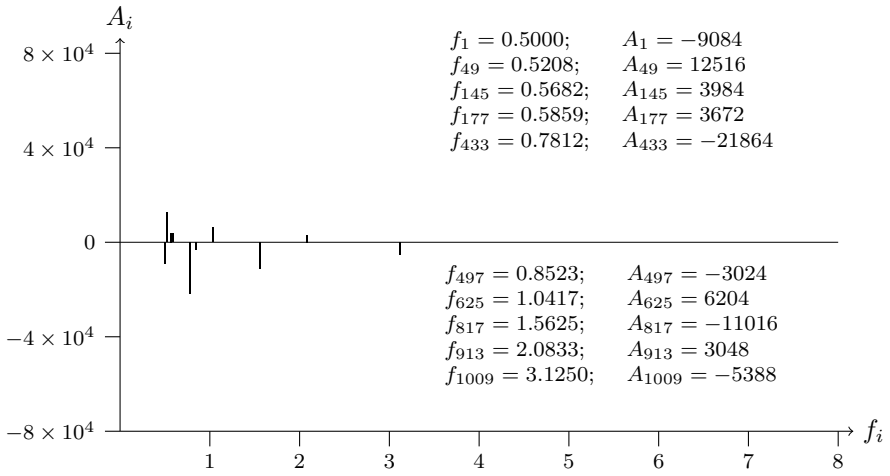

Figure 49a: T2T-CHM13v2.0; chromosome 12; VEZT exon-NM_001352092.2-13 analyzed data from 95280935 to 95282135

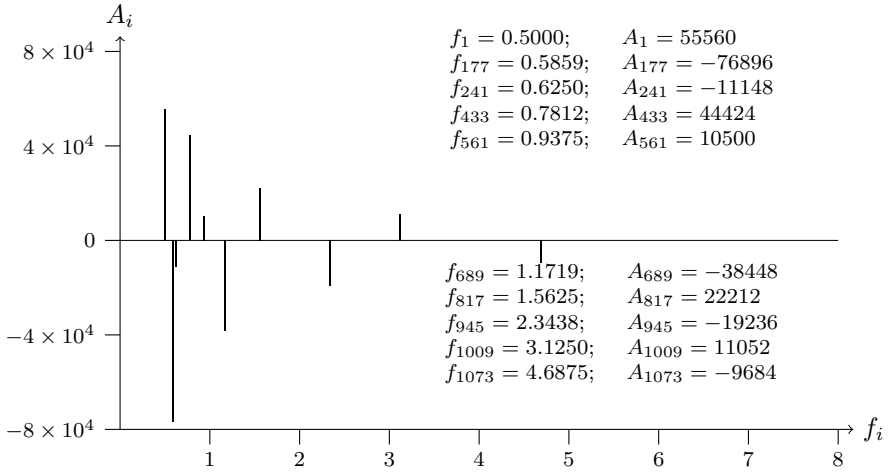

Figure 49b: T2T-CHM13v2.0; chromosome 12; VEZT exon-NM_001352092.2-13 analyzed data from 95282135 to 95280935

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

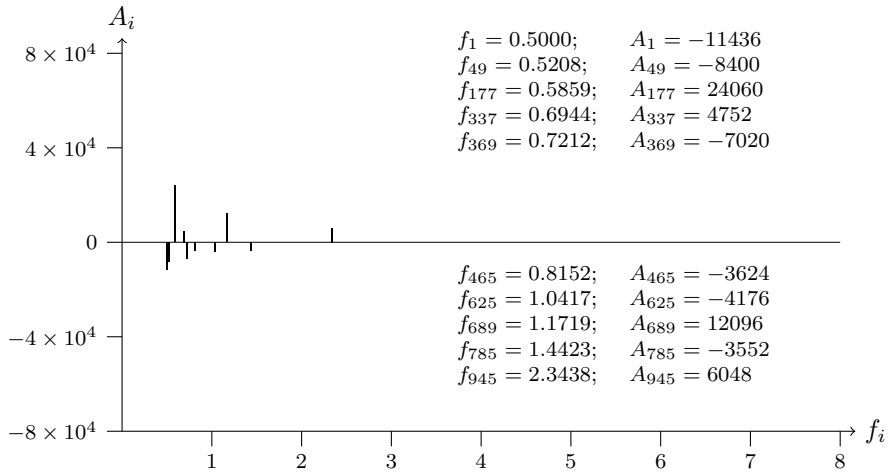

Figure 50a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-1 analyzed data from 95199140 to 95200340

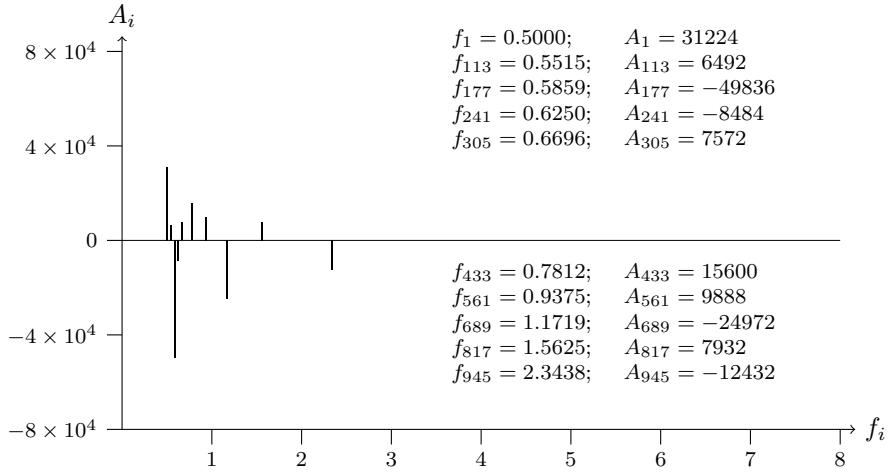

Figure 50b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-1 analyzed data from 95200340 to 95199140

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

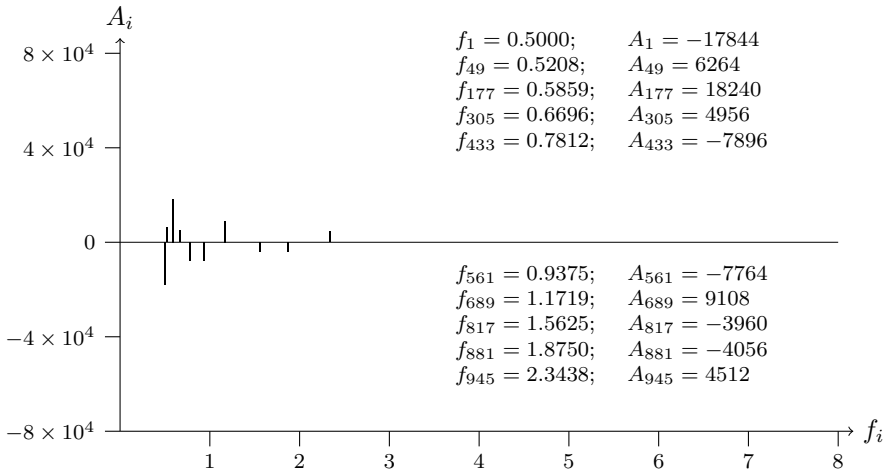

Figure 51a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-2 analyzed data from 95232839 to 95234039

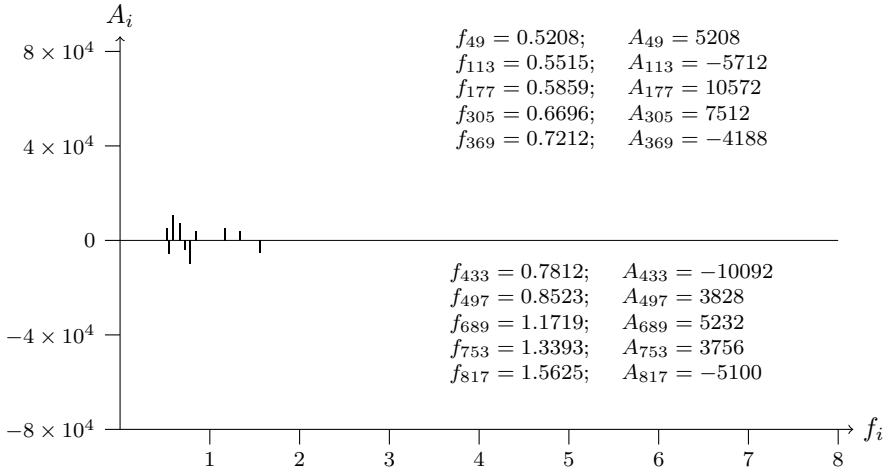

Figure 51b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-2 analyzed data from 95234039 to 95232839

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

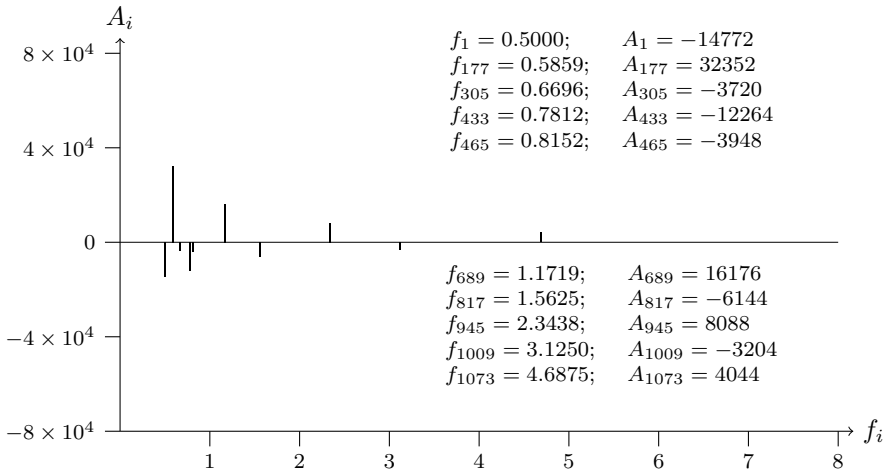

Figure 52a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-4 analyzed data from 95238006 to 95239206

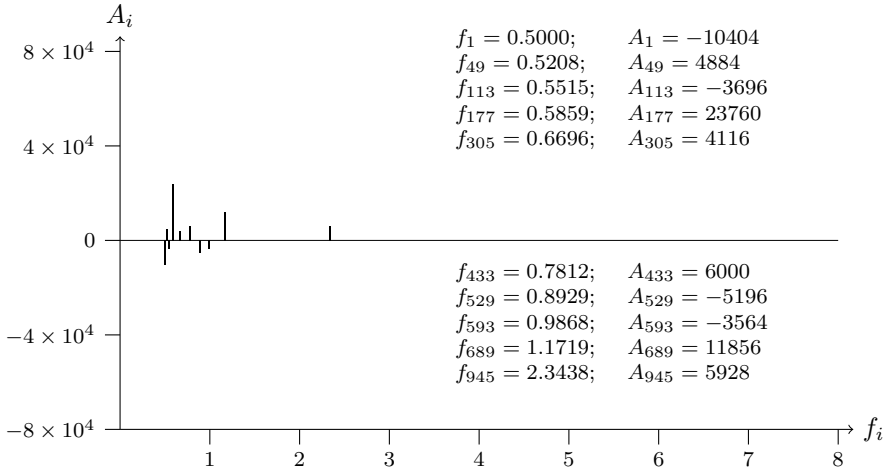

Figure 52b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-4 analyzed data from 95239206 to 95238006

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

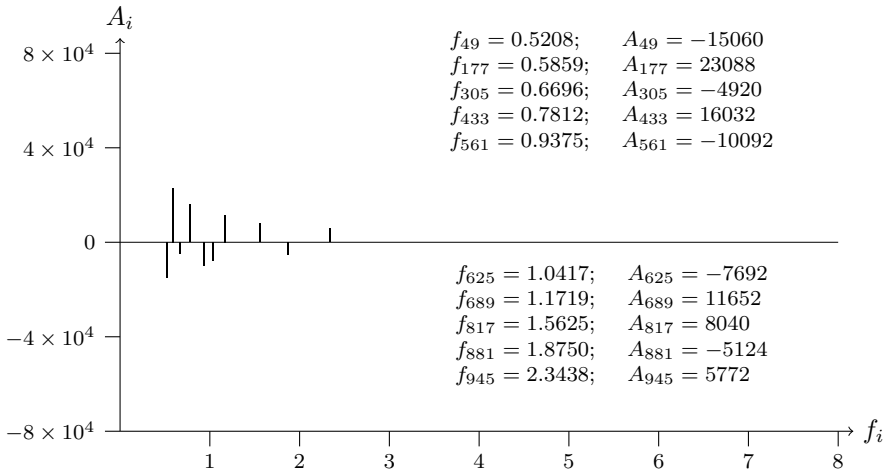

Figure 53a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-5 analyzed data from 95243850 to 95245050

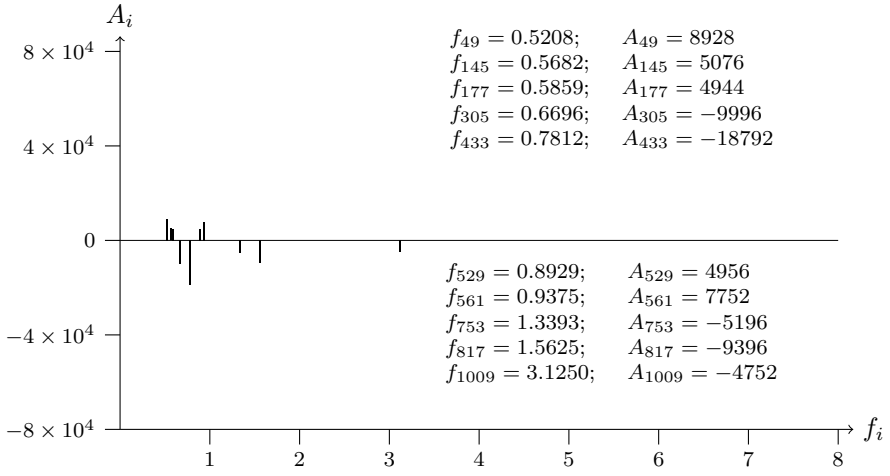

Figure 53b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-5 analyzed data from 95245050 to 95243850

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

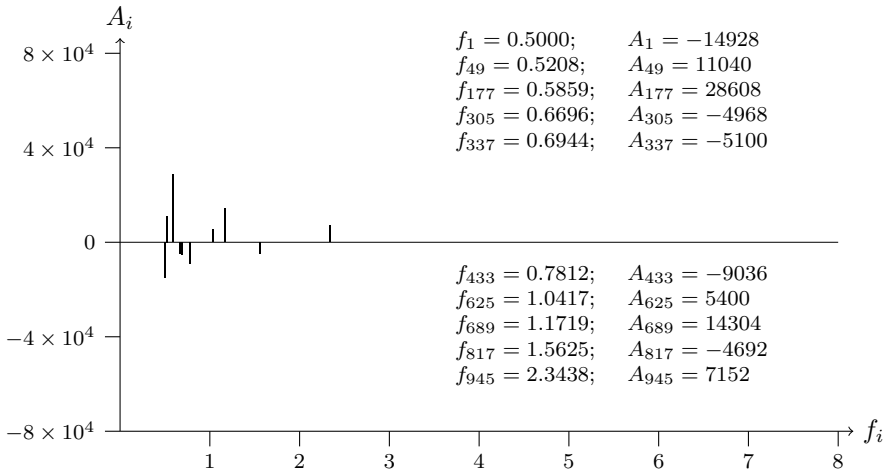

Figure 54a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-6 analyzed data from 95247401 to 95248601

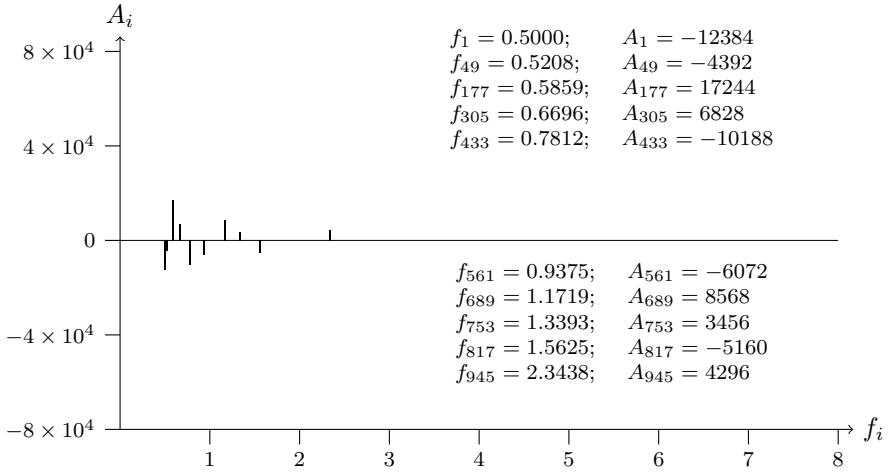

Figure 54b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-6 analyzed data from 95248601 to 95247401

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

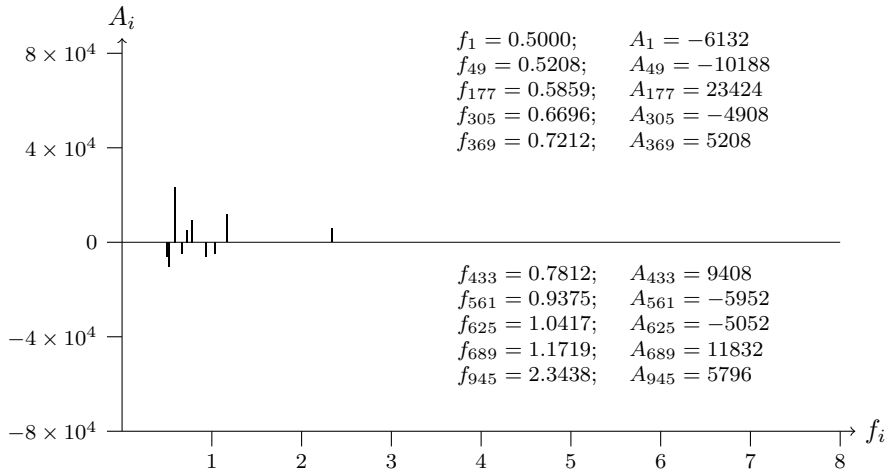

Figure 55a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-7 analyzed data from 95250957 to 95252157

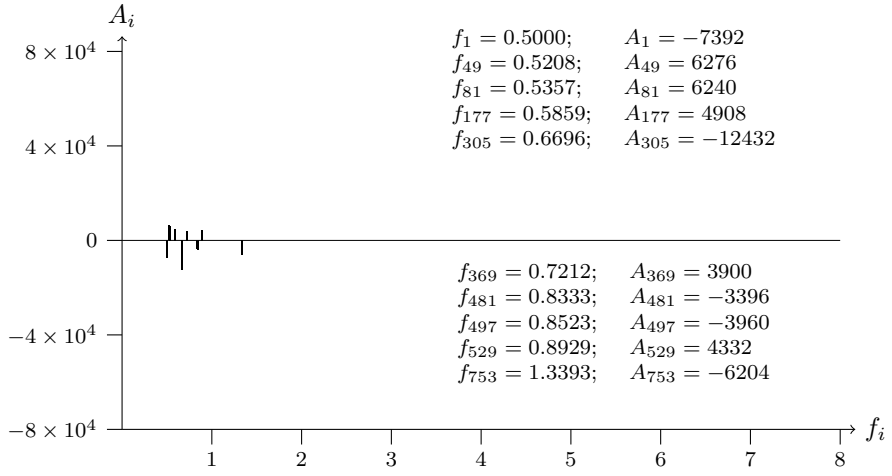

Figure 55b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-7 analyzed data from 95252157 to 95250957

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

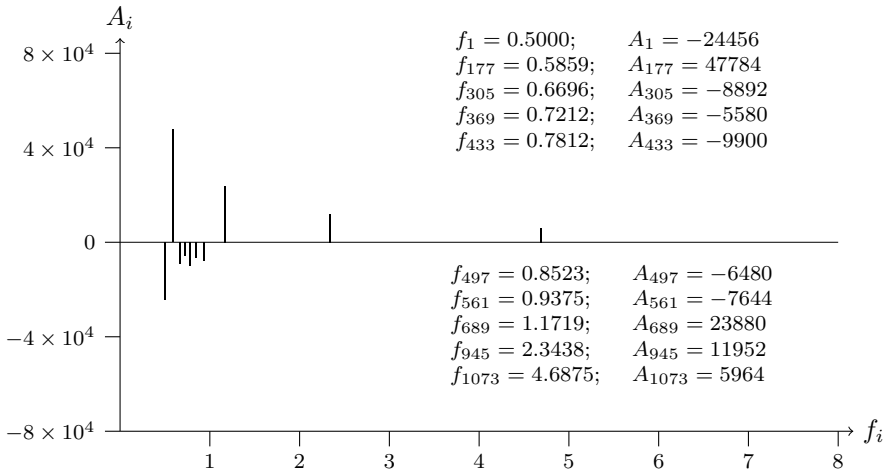

Figure 56a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-8 analyzed data from 95255658 to 95256858

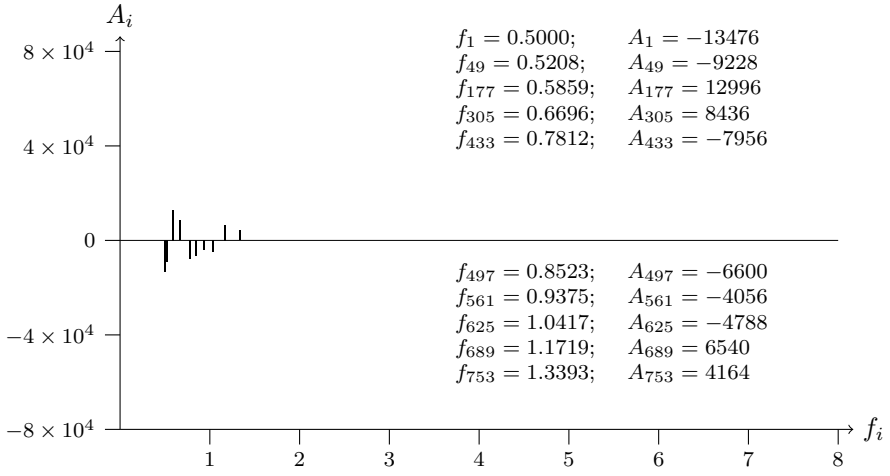

Figure 56b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-8 analyzed data from 95256858 to 95255658

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

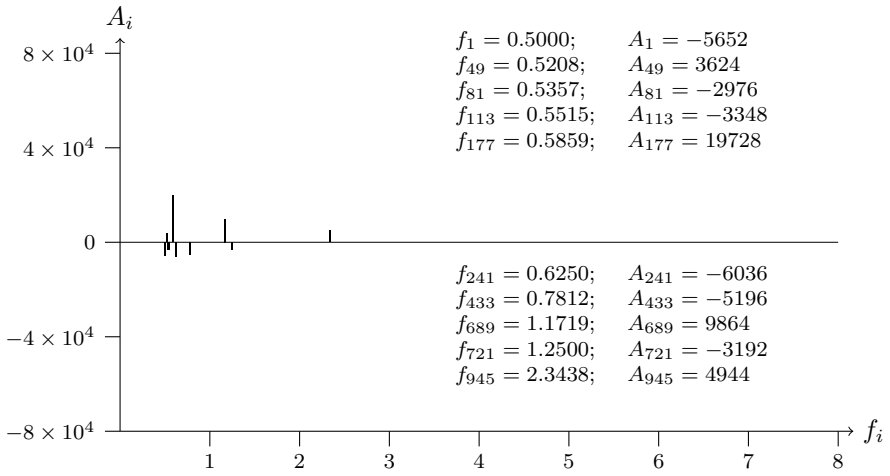

Figure 57a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-9 analyzed data from 95263413 to 95264613

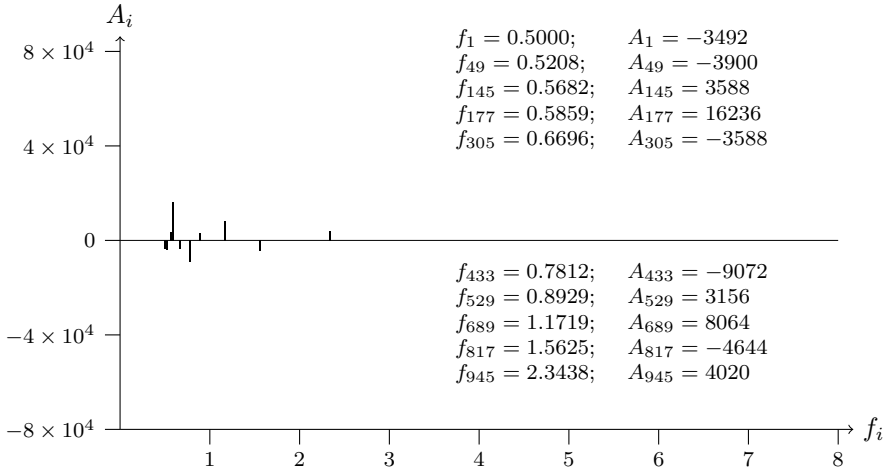

Figure 57b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-9 analyzed data from 95264613 to 95263413

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

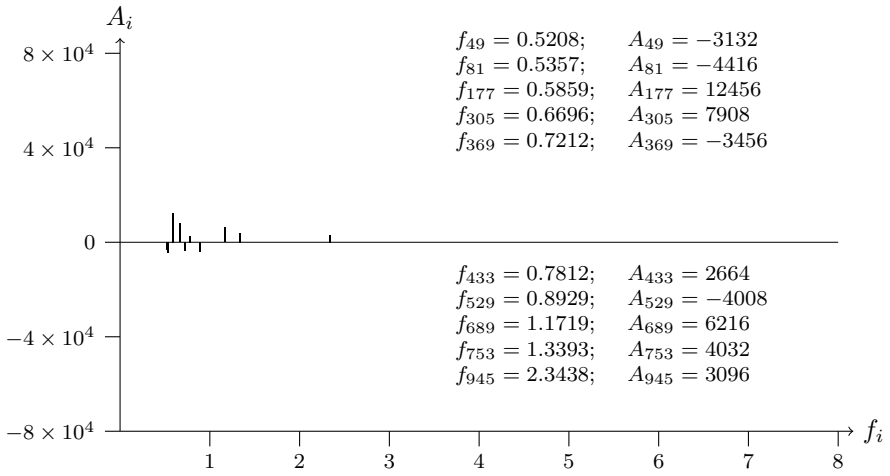

Figure 58a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-10 analyzed data from 95268628 to 95269828

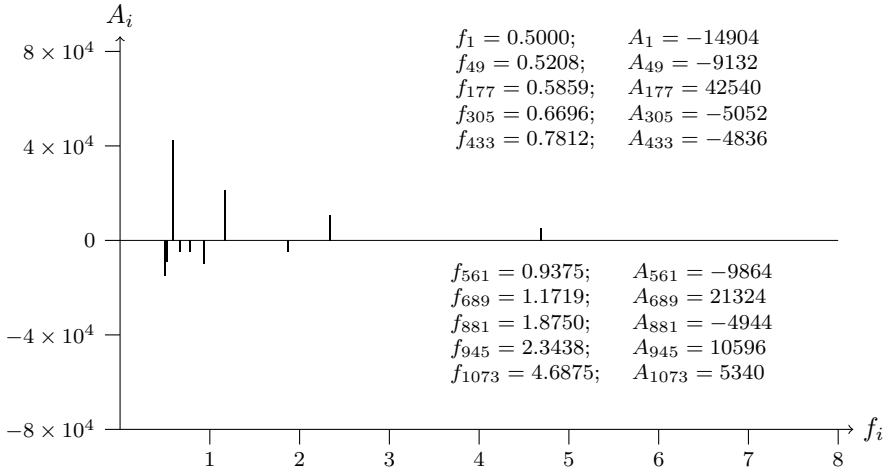

Figure 58b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-10 analyzed data from 95269828 to 95268628

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

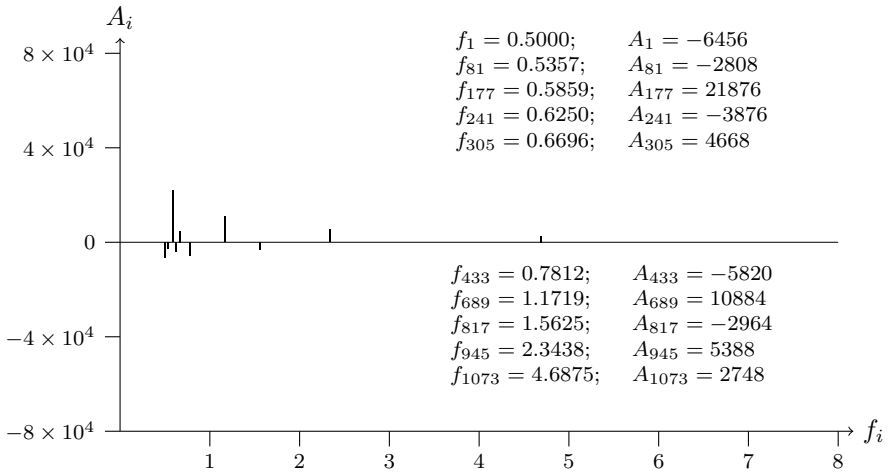

Figure 59a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-11 analyzed data from 95275142 to 95276342

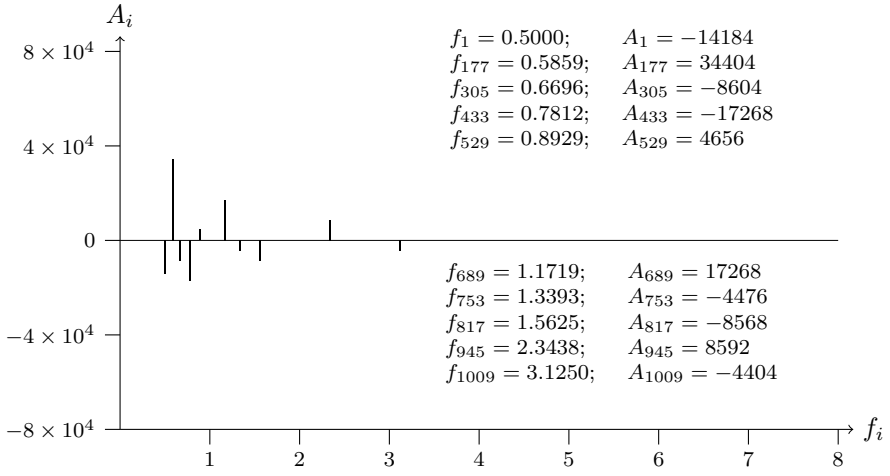

Figure 59b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-11 analyzed data from 95276342 to 95275142

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

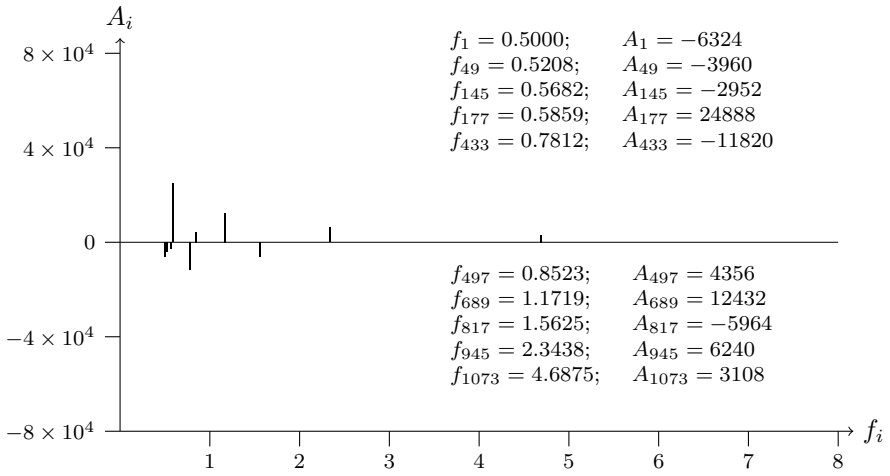

Figure 60a: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-12 analyzed data from 95277028 to 95278228

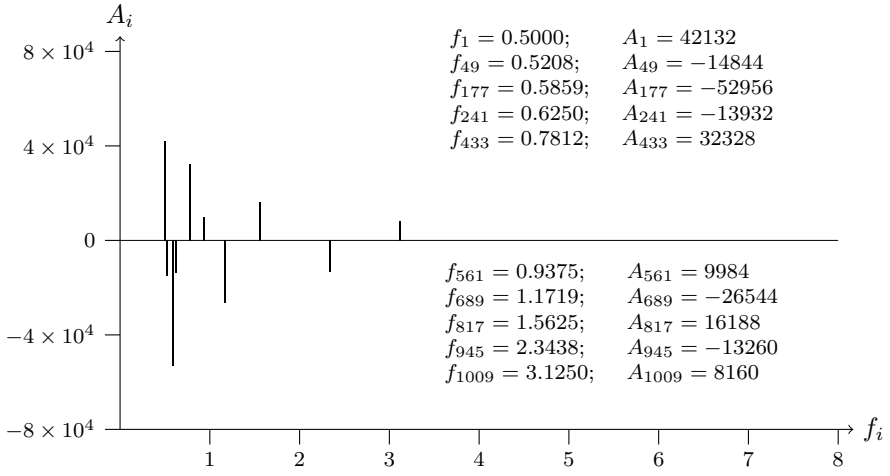

Figure 60b: T2T-CHM13v2.0; chromosome 12; VEZT intron-NM_001352092.2-12 analyzed data from 95278228 to 95277028

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

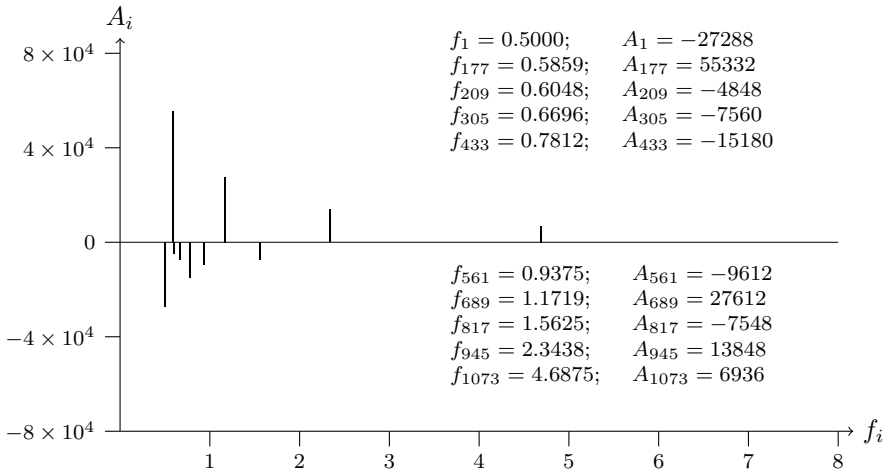

Figure 61a: T2T-CHM13v2.0; chromosome 12; TAOK3 exon-NM_016281.4-21 analyzed data from 118137047 to 118138247

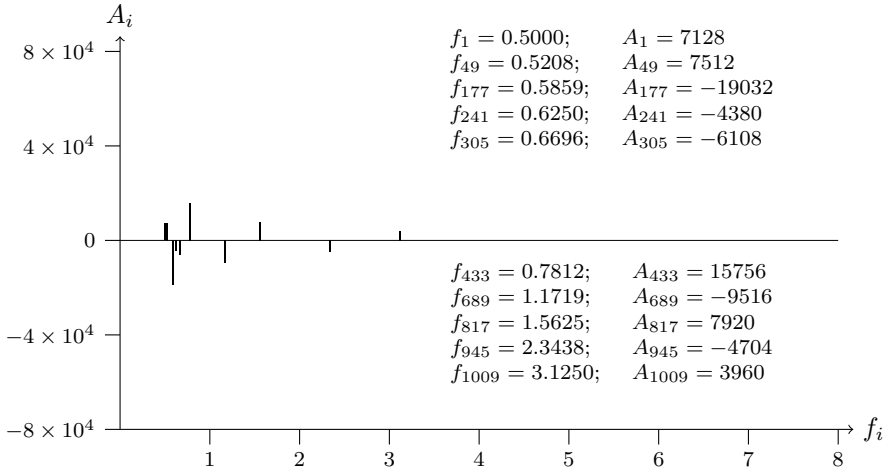

Figure 61b: T2T-CHM13v2.0; chromosome 12; TAOK3 exon-NM_016281.4-21 analyzed data from 118138247 to 118137047

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

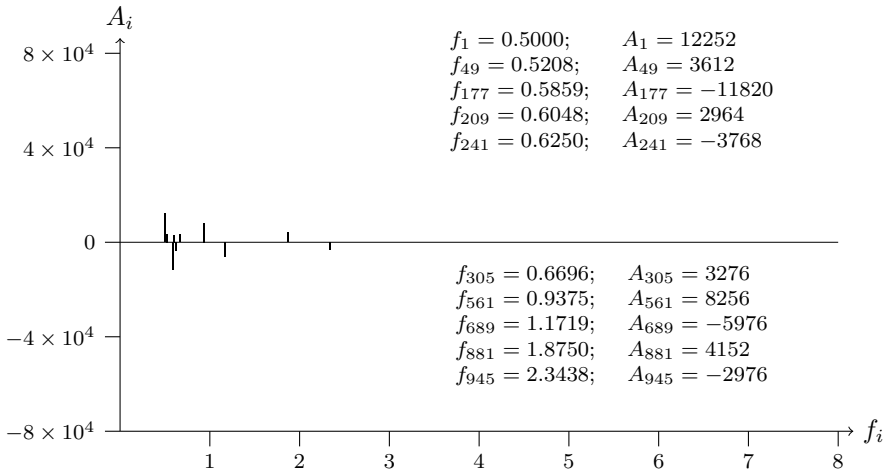

Figure 62a: T2T-CHM13v2.0; chromosome 12; SMUG1 exon-NM_001351259.2-2 analyzed data from 54145927 to 54147127

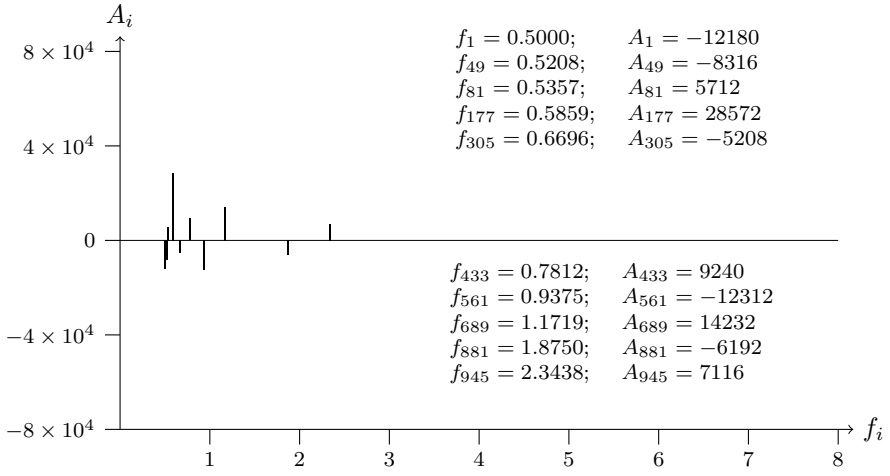

Figure 62b: T2T-CHM13v2.0; chromosome 12; SMUG1 exon-NM_001351259.2-2 analyzed data from 54147127 to 54145927

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

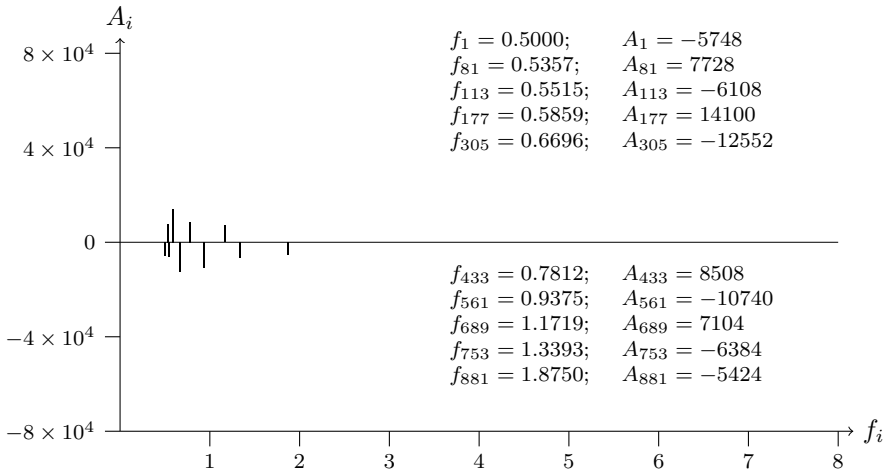

Figure 63a: T2T-CHM13v2.0; chromosome 4; MXD4 exon-NM_006454.3-6 analyzed data from 2246204 to 2247404

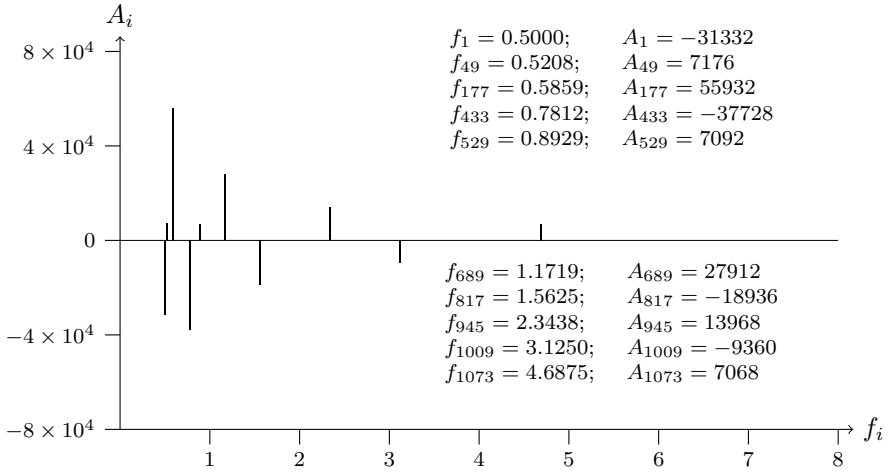

Figure 63b: T2T-CHM13v2.0; chromosome 4; MXD4 exon-NM_006454.3-6 analyzed data from 2247404 to 2246204

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

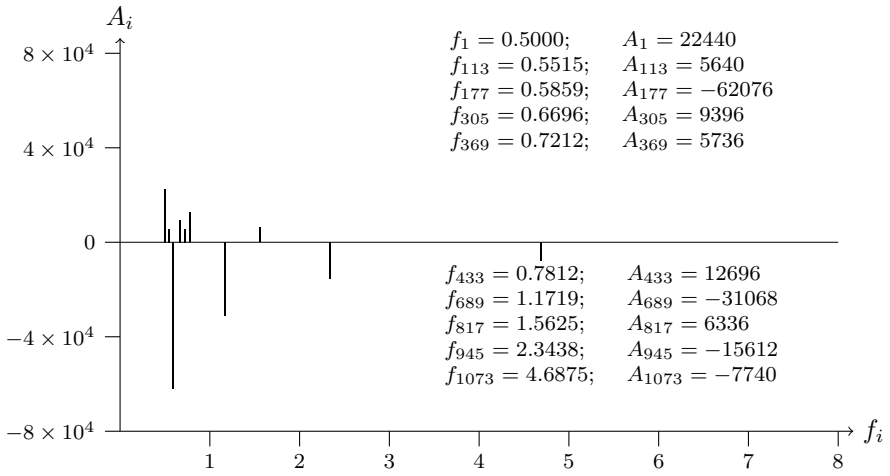

Figure 64a: T2T-CHM13v2.0; chromosome 4; AFAP1 exon-NM_001371091.1-18 analyzed data from 7730695 to 7731895

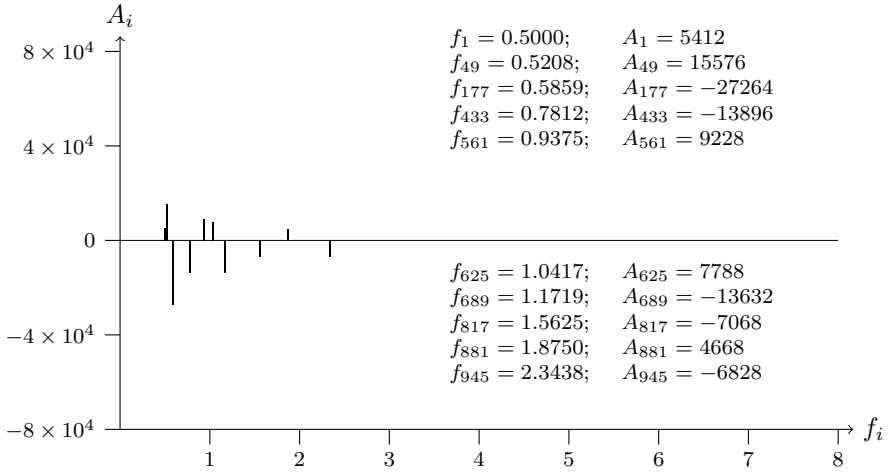

Figure 64b: T2T-CHM13v2.0; chromosome 4; AFAP1 exon-NM_001371091.1-18 analyzed data from 7731895 to 7730695

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

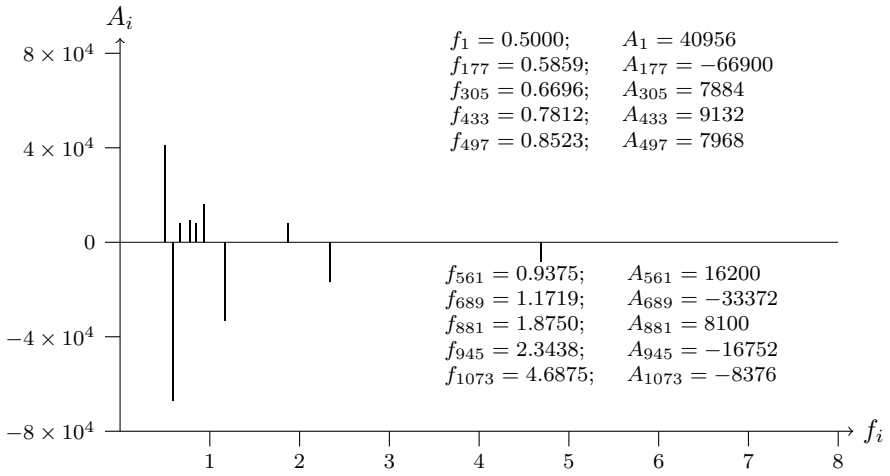

Figure 65a: T2T-CHM13v2.0; chromosome 19; SIGLEC10 exon-NM_001171161.2-9 analyzed data from 54498651 to 54499851

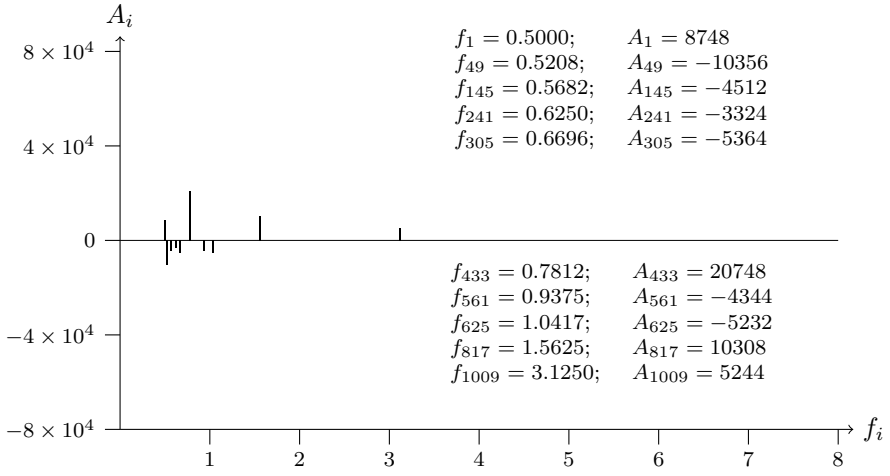

Figure 65b: T2T-CHM13v2.0; chromosome 19; SIGLEC10 exon-NM_001171161.2-9 analyzed data from 54499851 to 54498651

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

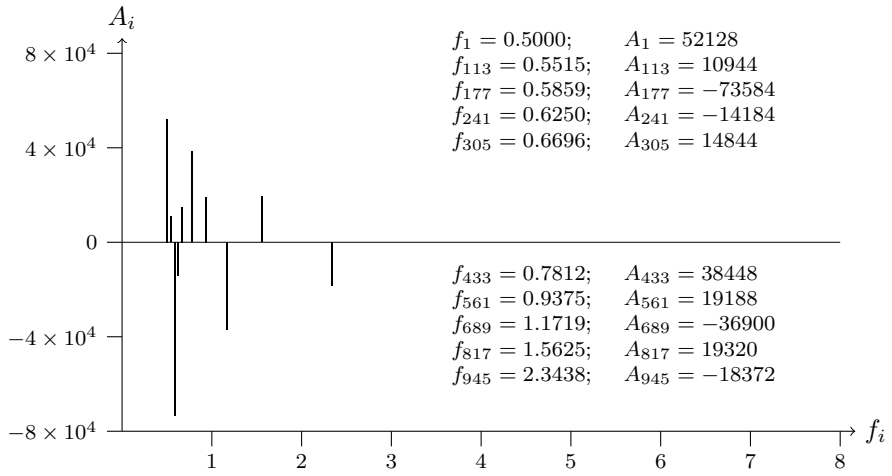

Figure 66a: T2T-CHM13v2.0; chromosome 19; CEACAM1 exon-NM_001184816.2-7 analyzed data from 45326710 to 45327910

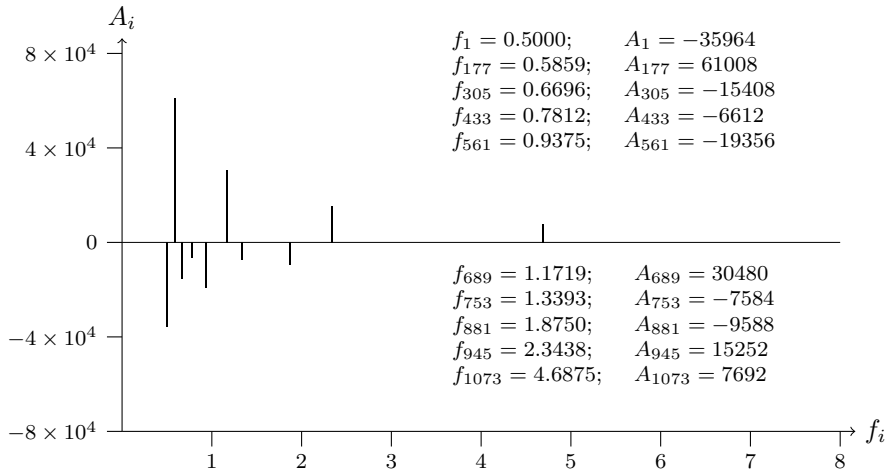

Figure 66b: T2T-CHM13v2.0; chromosome 19; CEACAM1 exon-NM_001184816.2-7 analyzed data from 45327910 to 45326710

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

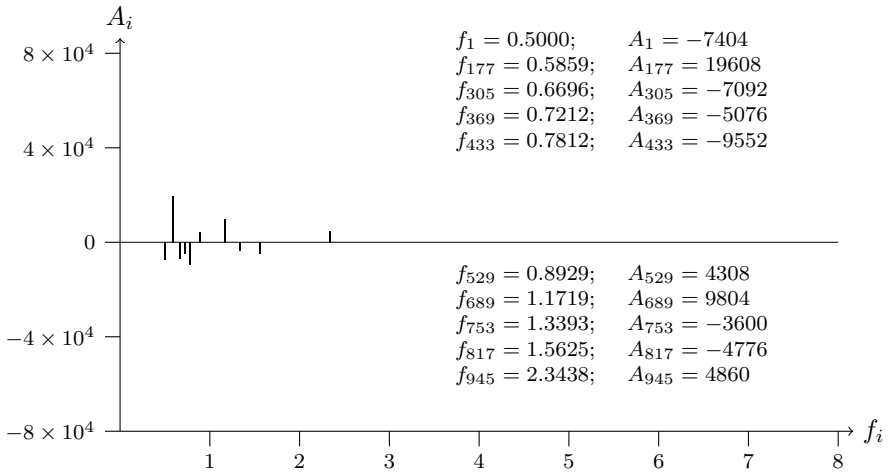

Figure 67a: T2T-CHM13v2.0; chromosome 16; ADGRG1 exon-NM_001370440.1-15 analyzed data from 63458652 to 63459852

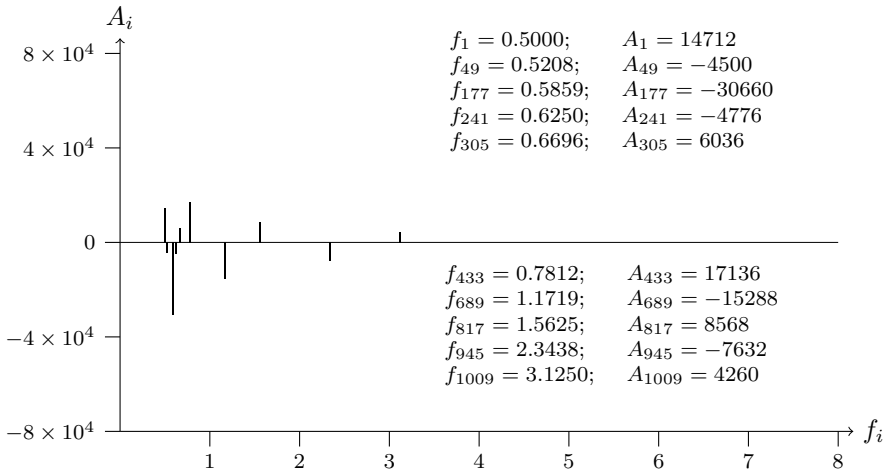

Figure 67b: T2T-CHM13v2.0; chromosome 16; ADGRG1 exon-NM_001370440.1-15 analyzed data from 63459852 to 63458652

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

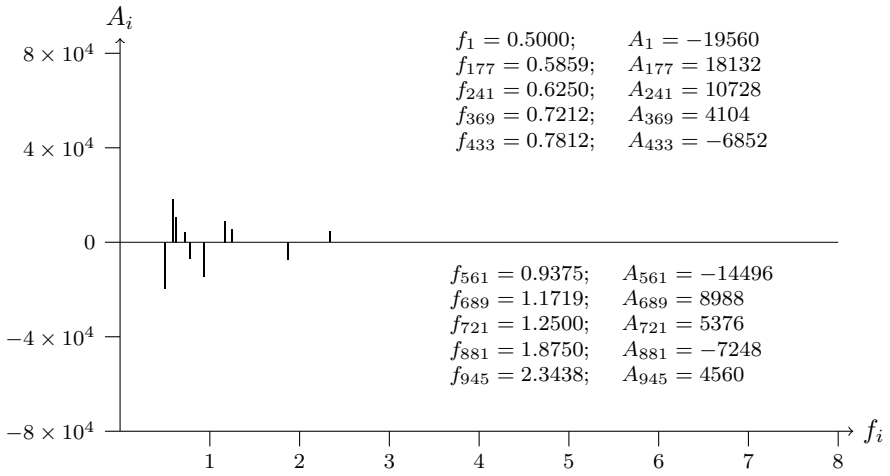

Figure 68a: T2T-CHM13v2.0; chromosome 16; KNOP1 exon-NM_001348534.2-4 analyzed data from 19632739 to 19633939

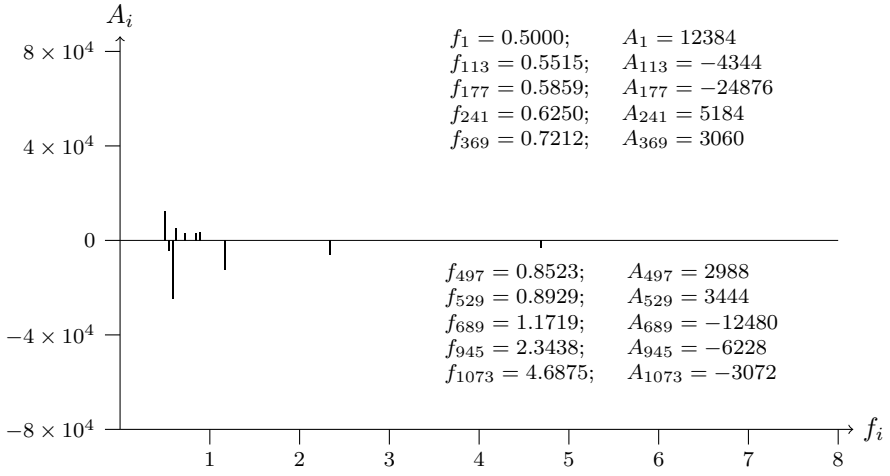

Figure 68b: T2T-CHM13v2.0; chromosome 16; KNOP1 exon-NM_001348534.2-4 analyzed data from 19633939 to 19632739

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

5 Preliminary Observations on Some of the Results Obtained

If instead of considering the values of the A_i and the f_i corresponding to the 10 greatest amplitudes of the 1200 A_i, for $i = 1, 2, 3, \ldots, 1200$, consideration is given to the A_i and the f_i corresponding to the 40 greatest amplitudes of those A_i, then the 40 A_i and the corresponding f_i will be known as the "prominent" ones. This is what was done when the following preliminary observations were carried out on some of the results obtained.

Attention was given to 77 samples from diverse parts of the human genome [131] (T2T-CHM13v2.0, downloaded from the National Center for Biotechnology Information Genome Database) of 1200 bases of DNA material [131]. Each sample of 1200 bases was analyzed in the same order as it is provided in the downloaded material [132].

Only the 40 most "prominent" A_i and their respective frequencies f_i were considered for each sample.

The information about the names and locations of genes, exons and introns has been extracted from secondary databases from the same datasource [132] (using the alignment file provided by the server). The listing of cancer-related genes (oncogenes and tumor suppressors) has also been obtained from NCBI [133].

Therefore, regarding the most common frequencies corresponding to those "prominent" A_i of the 77 samples, we can observe that:

1. Certain frequencies are present in 100% of the 77 samples, such as $f_{177} = 0.585938$ and $f_{689} = 1.171875$. This means that the presence of "prominent" A_i for these frequencies is related to genomic data in general.

2. Other frequencies are present in most cases (with no regard to any classification); for example, $f_{433} = 0.781250$ is "prominent" in 80% of the 77 cases analyzed.

3. Some frequencies distinguish specific subsets of cases:

a. $f_1 = 0.5$ is "prominent" in 80% of the telomeres analyzed and it is always "prominent" (100%) in the remaining samples.

b. $f_{241} = 0.625$ is closely related to exons, because it is "prominent" in 80% of the exons, with no distinction as to the type of exon (whether it is a normal exon or a cancer-related exon).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

c. $f_{625} = 1.041667$ is closely related to oncogenes and tumor suppression genes, since it is "prominent" in 80% of the exons and introns related to those genes in the subset analyzed.

d. $f_{81} = 0.535714$ and $f_{137} = 9.375$ can be interpreted as markers for introns, due to their appearance in 80% of the introns (with no distinction as to the type of the corresponding gene).

e. $f_{305} = 0.669634$ is not present in telomeres; thus it can be used to distinguish between telomeric data and non-telomeric data.

f. $f_{817} = 1.5625$ and $f_{945} = 2.34375$ appear to group non-coding data (telomeres, centromeres, introns); that occurs in 90% of samples analyzed.

g. $f_{369} = 0.721154$ is "prominent" in 80% of the samples of telomeres and introns, but that frequency does not appear to be significant in centromeres.

h. $f_{49} = 0.520833$ is present in all the cancer-related exons (these exons are part of oncogenes and tumor suppressor genes); and in 80% of the introns and telomeres.

6 Discussion and Perspectives

The SWM has certain characteristics of interest with respect to other possible methods for the analysis of numerical sequences. Thus, for example, a) it can be applied both for overall and local analysis; b) it has no "effect" such as that of the Gibbs phenomenon which would require corrections to ensure the quality of the approximations achieved to the numerical sequences analyzed; c) it depends on the order in which the elements constituting each numerical sequence analyzed are presented, and not simply on the statistical distribution of those elements; and d) the complete results of each analysis make it possible to reconstruct, unambiguously, the sequence of bases of DNA analyzed.

The preliminary observations specified in section 4 suggest certain patterns that are detected by the SWM. Experts in genetics and bioinformatics will determine whether these interesting regularities support the hypothesis formulated in the introduction of this article – in which case using data analysis techniques to go into greater depth on the significance of the results obtained would be justified – or whether this hypothesis should be discarded.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

[1] Skliar, O., Monge, R. E., & Gapper, S. (2016). Analysis of time series and signals using the Square Wave Method. arXiv preprint *arXiv*:1608.02166 [cs.NA].

[2] Skliar, O., Monge, R. E., & Gapper, S. (2015). A new method for signal and image analysis: The Square Wave Method. arXiv preprint *arXiv*:1501.0068 [cs.NA].

[3] Skliar, O., Monge, R. E., Oviedo, G., & Gapper, S. (2016). A new method for the analysis of signals: The Square Wave Transform. *Revista de Matemática: Teoría y Aplicaciones*, Vol. 23(1), pp. 85-110.

[4] Skliar, O., Oviedo, G.,Monge, R. E., Medina, V., & Gapper, S. (2013). A new method for the analysis of images: The Square Wave Method. *Revista de Matemática: Teoría y Aplicaciones*, Vol. 20(2), pp. 133-153.

[5] Skliar, O., Medina, V., & Monge, R. E. (2008). A new method for the analysis of signals: The Square Wave Method. *Revista de Matemática: Teoría y Aplicaciones*, Vol. 15(2), pp. 109-129.

[6] Wiebe, N., & Roetteler, M. (2014). Quantum arithmetic and numerical analysis using Repeat-Until-Success circuits. Redmond, WA: Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA. arXiv preprint *arXiv*:1406.2040, pp. 1-32 (22, 32).

[7] Wiebe, N., & Roetteler, M. (2016). Quantum arithmetic and numerical analysis using Repeat-Until-Success circuits. *Quantum Information & Computation*, Vol. 16(1-2), pp. 134-178.

[8] Kivi, J. (2019). Algorithmic road state modelling/Algoritminen tienpinnan tilan mallinnus. Thesis. Espoo, FI: Aalto University/Aalto-yliopisto, pp. 36, 55.

[9] Aslam, A. (2023). Universal Language Modelling agent – ULMa, https://arxiv.org/ftp/arxiv/papers/2306/2306.06521.pdf. pp. 1-12 (6, 11).

[10] Olanrewaju, R. O., Olanrewaju, S. A., Folorunsho, S., & Gbemisola Dada, A. G. (2023). Time variant wave-signal-amplitude trigonometry regression of latitudes and longitudes of the Belmullets of the Atlantic Ocean. *European Journal of Statistics*, Vol. 3(12), pp. 1-12 (2, 11). https://doi.org/10.28924/ada/stat.3.12.

[11] Fonkou, R. F., Kengne, R., Kamgang, H. C. F., & Talla, P. K. (2023). Dynamical behavior analysis of the Van der Pol oscillator with sine nonlinearity subjected to non-sinusoidal periodic excitations by the bifurcation structures. *Physica Scripta* 98, 085014, pp. 3, 20. https://doi.org/10.1088/1402-4896/ace8cf.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[12] Panta Vásquez, L. M. (2016). Diseño e implementación de redes HFC para la ampliación del servicio 3Play. Thesis, Universidad Nacional Tecnológica de Lima Sur, pp. 9, 50.

[13] Amiri, O. T. & Simpson, R., Kandou Labs SA (2018). Low power multi-level driver. U.S. Patent 10, 056,903. (Skliar et al., 2008, cited in "Other publications", p. 5).

[14] Amiri, O. T. & Tajalli A., Kandou Labs SA (2019). Phase interpolation circuit. U.S. Patent 10, 203,226. (Skliar et al., 2008, cited in "Other publications", p. 5).

[15] Cronie, H., & Holden, B., Kandou Labs SA (2016). Sorting decoder. U.S. Patent 9,362,947. (Skliar et al., 2008, cited in "Other publications", p. 3).

[16] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Differential vector storage for dynamic random access memory. U.S. Patent 9,275,720. (Skliar et al., 2008, cited in "Other publications", p. 3).

[17] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Differential vector storage for dynamic random access memory. U.S. Patent 9,424,908. (Skliar et al., 2008, cited in "Other publications", p. 3).

[18] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2017). Error control coding for the orthogonal differential vector signaling. U.S. Patent 9,667,379. (Skliar et al., 2008, cited in "Other publications", p. 3).

[19] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Methods and systems for low power and pin-efficient communications with superposition signaling coldes. U.S. Patent 9,401,828. (Skliar et al., 2008, cited in "Other publications", p. 3).

[20] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Orthogonal differential vector signaling. U.S. Patent 9,288,089. (Skliar et al., 2008, cited in "Other publications", p. 3).

[21] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2017). Orthogonal differential vector signaling. U.S. Patent 9,825,677. (Skliar et al., 2008, cited in "Other publications", p. 4).

[22] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2019). Orthogonal differential vector signaling. U.S. Patent 10,355,756. (Skliar et al., 2008, cited in "Other publications", p. 4).

[23] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2021). Orthogonal differential vector signaling. U.S. Patent 10,985,806. (Skliar et al., 2008, cited in "Other publications", p. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[24] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2023). Orthogonal differential vector signaling. U.S. Patent 11,611,377 (Skliar et al., 2008, cited in "Other Publications", p. 3).

[25] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Storage method and apparatus for random access memory codeword storage. U.S. Patent 9,268,683. (Skliar et al., 2008, cited in "Other publications", p. 3).

[26] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Storage method and apparatus for random access memory codeword storage. U.S. Patent 9,524,106. (Skliar et al., 2008, cited in "Other publications", p. 4).

[27] Cronie, H., & Shokrollahi, A., Ecole Polytechnique Federale de Lausanne EPFL (2016). Storage method and apparatus for random access memory codeword storage. U.S. Patent 9,361,223. (Skliar et al., 2008, cited in "Other publications", p. 3).

[28] Fox, J., Holden, B., Hunt, P., Keay, J. D., Shokrollahi, A., Simpson, R., Singh, A., Kevin, A., Stewart, J., Surace, G., Kandou Labs SA (2016). Methods and systems for chip-to-chip communication with reduced simultaneous switching noise. U.S. Patent 9,357,036. (Skliar et al., 2008, cited in "Other publications", p. 3).

[29] Fox, J., Holden, B., Hunt, P., Keay, J. D., Shokrollahi, A., Simpson, R., Singh, A., Kevin, A., Stewart, J., Surace, G., Kandou Labs SA (2017). Methods and systems for chip-to-chip communication with reduced simultaneous switching noise. U.S. Patent 9,686,107. (Skliar et al., 2008, cited in "Other publications", p. 4).

[30] Fox, J., Holden, B., Hunt, P. Keay, J. D., Shokrollahi, A., Kevin, A., Stewart, J., Surace, G., Ulrich, R., & Simpson, R., Kandou Labs SA (2016). Methods and systems for high bandwidth chip-to-chip communications interface. U.S. Patent 9,362,974. (Skliar et al., 2008, cited in "Other publications", p. 3).

[31] Fox, J., Holden, B., Hormati, A., Hunt, P., Keay, J. D., Shokrollahi, A., Simpson, R., Kevin, A., Stewart, J., Surace, G., & Ulrich, R., Kandou Labs SA (2017). Methods and systems for high bandwidth communications interface. U.S. Patent 9,596,109. (Skliar et al., 2008, cited in "Other publications", p. 3).

[32] Fox, J., Holden, B., Hunt, P. Keay, J. D., Shokrollahi, A., Kevin, A., Stewart, J., Surace, G., Ulrich, R., & Simpson, R., Kandou Labs SA (2017). Methods and systems for high bandwidth chip-to-chip communications interface. U.S. Patent 9,838,017. (Skliar et al., 2008, cited in "Other publications", p. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[33] Fox, J., Holden, B., Hormati, A., Hunt, P., Keay, J. D., Shokrollahi, A., Simpson, R., Kevin, A., Stewart, J., Surace, G., & Ulrich, R., Kandou Labs SA (2018). Methods and systems for high bandwidth communications interface. U.S. Patent 10,091,135. (Skliar et al., 2008, cited in "Other publications", p. 3).

[34] Fox, J., Holden, B., Hormati, A., Hunt, P., Keay, J. D., Shokrollahi, A., Simpson, R., Kevin, A., Stewart, J., Surace, G., & Ulrich, R., Kandou Labs SA (2022). Methods and systems for high bandwidth communications interface. U.S. Patent 11,374, 801. (Skliar et al., 2008, cited in "Other publications", p. 5).

[35] Fox, J., Holden, B., Shokrollahi, A., Singh, A., & Surace, G., Kandou Labs SA (2016). Methods and systems for pin-efficient memory controller interface using vector signaling codes for chip-to-chip communications. U.S. Patent 9,251,873. (Skliar et al., 2008, cited in "Other publications", p. 3).

[36] Fox, J., Holden, B., Shokrollahi, A., Singh, A., & Surace, G., Kandou Labs SA (2017). Methods and systems for pin-efficient memory controller interface using vector signaling codes for chip-to-chip communications. U.S. Patent 9,607,673. (Skliar et al., 2008, cited in "Other publications", p. 4).

[37] Fox, J., Holden, B., Shokrollahi, A., Singh, A., & Surace, G., Kandou Labs SA (2016). Methods and systems for pin-efficient memory controller interface using vector signaling codes for chip-to-chip communications. U.S. Patent 10,468,078. (Skliar et al., 2008, cited in "Other publications", p. 6).

[38] Holden, B., & Shokrollahi, A., Kandou Labs SA (2018). Methods and systems for selection of unions of vector signaling codes for power and pin-efficient chip-to-chip communication. U.S. Patent 9,929,818. (Skliar et al., 2008, cited in "Other publications", p. 4).

[39] Holden, B., & Shokrollahi, A., Kandou Labs SA (2016). Orthogonal differential vector signaling codes with embedded clock. U.S. Patent 9,461,862. (Skliar et al., 2008, cited in "Other publications", p. 3).

[40] Holden, B., & Shokrollahi, A., Kandou Labs SA (2017). Orthogonal differential vector signaling codes with embedded clock. U.S. Patent 9,838,234. (Skliar et al., 2008, cited in "Other publications", p. 5).

[41] Holden, B., & Shokrollahi, A., Kandou Labs SA (2018). Orthogonal differential vector signaling codes with embedded clock. U.S. Patent 10,122,561. (Skliar et al., 2008, cited in "Other publications", p. 5).

[42] Holden, B., & Shokrollahi, A., Kandou Labs SA (2020). Orthogonal differential vector signaling codes with embedded clock. U.S. Patent 10,652,067. (Skliar et al., 2008, cited in "Other publications", p. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[43] Holden, B., Shokrollahi, A., & Singh, A., Kandou Labs SA (2016). Methods and systems for skew tolerance in and advanced detectors for vector signaling codes for chip-to-chip communication. U.S. Patent 9,300,503. (Skliar et al., 2008, cited in "Other publications", p. 2).

[44] Holden, B., Shokrollahi, A., & Singh, A., Kandou Labs SA (2017). Methods and systems for skew tolerance in and advanced detectors for vector signaling codes for chip-to-chip communication. U.S. Patent 9,825,723. (Skliar et al., 2008, cited in "Other publications", p. 4).

[45] Holden, B., Shokrollahi, A., & Singh, A., Kandou Labs SA (2018). Methods and systems for skew tolerance in and advanced detectors for vector signaling codes for chip-to-chip communication. U.S. Patent 10,044,452. (Skliar et al., 2008, cited in "Other publications", p. 5).

[46] Hormati, A., Kandou Labs SA (2018). Low power chip-to-chip bidirectional communications. U.S. Patent 10,116,,468. (Skliar et al., 2008, cited in "Other publications", p. 5).

[47] Hormati, A., Kandou Labs SA (2020). Low power chip-to-chip bidirectional communications. U.S. Patent 10,581,644. (Skliar et al., 2008, cited in "Other publications", p. 3).

[48] Hormati, A., Kandou Labs SA (2021). Low power chip-to-chip bidirectional communications. U.S. Patent 11,032,110. (Skliar et al., 2008, cited in "Other publications", p. 3).

[49] Hormati, A., Kandou Labs SA (2022). Low power chip-to-chip bidirectional communications. U.S. Patent 11,477,055. (Skliar et al., 2008, cited in "Other publications", p. 3).

[50] Hormati, A., Kandou Labs SA (2016). Multidrop data transfer. U.S. Patent 9,444,654. (Skliar et al., 2008, cited in "Other publications", p. 3).

[51] Hormati, A., Kandou Labs SA (2018). Multidrop data transfer. U.S. Patent 9,893,911. (Skliar et al., 2008, cited in "Other publications", p. 5).

[52] Hormati, A., & Shokrollahi, A., Kandou Labs SA (2019). Multidrop data transfer. U.S. Patent 10,230,549. (Skliar et al., 2008, cited in "Other publications", p. 5).

[53] Hormati, A., Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2016). Method and apparatus for low power chip-to-chip communication with constrained ratio. U.S. Patent 9,258,154. (Skliar et al., 2008, cited in "Other publications", p. 2).

[54] Hormati, A., Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2017). High speed communications system. U.S. Patent 9,832,046. (Skliar et al., 2008, cited in "Other publications", p. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[55] Hormati, A., Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2018). High speed communications system. U.S. Patent 10,116,472. (Skliar et al., 2008, cited in "Other publications", p. 5).

[56] Hormati, A., Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2020). High speed communications system. U.S. Patent 10,608,850. (Skliar et al., 2008, cited in "Other publications", p. 3).

[57] Hormati, A., Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2020). High speed communications system. U.S. Patent 10,819,541. (Skliar et al., 2008, cited in "Other publications", p. 4).

[58] Hormati, A., Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2021). High speed communications system. U.S. Patent 11,115,249 (Skliar et al., 2008, cited in "Other publications", p. 4).

[59] Hormati, A., Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2022). High speed communications system. U.S. Patent 11,483,187 (Skliar et al., 2008, cited in "Other publications", p. 4).

[60] Shokrollahi, A. Kandou Labs (2017). Methods and systems for reduction of nearest-neighbor crosstalk. U.S. Patent 9,806,761. (Skliar et al., 2008, cited in "Other publications", p. 4).

[61] Shokrollahi, A. Kandou Labs (2019). Methods and systems for reduction of nearest-neighbor crosstalk. U.S. Patent 10,177,812. (Skliar et al., 2008, cited in "Other publications", p. 5).

[62] Shokrollahi, A. Kandou Labs (2018). Orthogonal differential vector signaling codes with embedded clock. U.S. Patent 10,055,372. (Skliar et al., 2008, cited in "Other publications", p. 5).

[63] Shokrollahi, A. Kandou Labs (2016). Vector signaling codes with high pin-efficiency for chip-to-chip communication and storage. U.S. Patent 9,479,369. (Skliar et al., 2008, cited in "Other publications", p. 3).

[64] Shokrollahi, A. Kandou Labs (2019). Vector signaling code with improved noise margin. U.S. Patent 10,333,749. (Skliar et al., 2008, cited in "Other publications", p. 5).

[65] Shokrollahi, A. Kandou Labs (2022). Vector signaling code with improved noise margin. U.S. Patent 11,240,076. (Skliar et al., 2008, cited in "Other publications", p. 3).

[66] Shokrollahi, A. Kandou Labs (2023). Vector signaling code with improved noise margin. U.S. Patent 11,716,227. (Skliar et al., 2008, cited in "Other publications", p. 3).

[67] Shokrollahi, A. Kandou Labs (2016). Vector signaling with reduced receiver complexity. U.S. Patent 9,246,713. (Skliar et al., 2008, cited in "Other publications", p. 2).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[68] Shokrollahi, A. Kandou Labs (2016). Vector signaling with reduced receiver complexity. U.S. Patent 9,485,057. (Skliar et al., 2008, cited in "Other publications", p. 3).

[69] Shokrollahi, A. Kandou Labs (2017). Vector signaling with reduced receiver complexity. U.S. Patent 9,692,555. (Skliar et al., 2008, cited in "Other publications", p. 5).

[70] Shokrollahi, A. Kandou Labs (2018). Vector signaling with reduced receiver complexity. U.S. Patent 9,985,745. (Skliar et al., 2008, cited in "Other publications", p. 5).

[71] Shokrollahi, A., & Carnelli, D., Kandou Labs SA (2020). Pipelined forward error correction for vector signaling code channel. U.S. Patent 10,666,297. (Skliar et al., 2008, cited in "Other publications", p. 4).

[72] Shokrollahi, A., & Carnelli, D., Kandou Labs SA (2022). Pipelined forward error correction for vector signaling code channel. U.S. Patent 11,336,302. (Skliar et al., 2008, cited in "Other publications", p. 4).

[73] Shokrollahi, A., & Carnelli, D., Kandou Labs SA (2023). Pipelined forward error correction for vector signaling code channel. U.S. Patent 11,804,845. (Skliar et al., 2008, cited in "Other publications", p. 4).

[74] Shokrollahi, A., & Chiurtu, N., Kandou Labs SA (2016). Low EMI signaling for parallel conductor interfaces. U.S. Patent 9,369,312. (Skliar et al., 2008, cited in "Other publications", p. 3).

[75] Shokrollahi, A., & Hormati, A., Kandou Labs SA (2020). Multi-wire permuted forward error correction. U.S. Patent 10,693,587. (Skliar et al., 2008, cited in "Other publications", p. 4).

[76] Shokrollahi, A., & Hormati, A., Kandou Labs SA (2022). Multi-wire permuted forward error correction. U.S. Patent 11,368,247. (Skliar et al., 2008, cited in "Other publications", p. 4).

[77] Shokrollahi, A., & Johnston, M. W., Kandou Labs SA (2018). Skew-resistant multi-wire channel. U.S. Patent 10,153,591. (Skliar et al., 2008, cited in "Other publications", p. 5).

[78] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2017). Fault tolerant chip-to-chip communication with advanced voltage. U.S. Patent 9,564,994. (Skliar et al., 2008, cited in "Other publications", p. 3).

[79] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2017). Method and apparatus for high speed chip-to-chip communications. U.S. Patent 9,674,014. (Skliar et al., 2008, cited in "Other Publications", p. 4).

[80] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2019). Method and apparatus for high speed chip-to-chip communications. U.S. Patent 10,243,765. (Skliar et al., 2008, cited in "Other Publications", p. 5).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[81] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2016). Vector signaling codes with increased signal to noise characteristics. U.S. Patent 9,509,437. (Skliar et al., 2008, cited in "Other Publications", p. 3).

[82] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2017). Vector signaling codes with increased signal to noise characteristics. U.S. Patent 9,900,186. (Skliar et al., 2008, cited in "Other Publications", p. 4).

[83] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2018). Vector signaling codes with increased signal to noise characteristics. U.S. Patent 10,020,966. (Skliar et al., 2008, cited in "Other Publications", p. 5).

[84] Shokrollahi, A., & Ulrich, R., Kandou Labs SA (2019). Vector signaling codes with increased signal to noise characteristics. U.S. Patent 10,320,588. (Skliar et al., 2008, cited in "Other Publications", p. 6).

[85] Shokrollahi, A., Holden, B., & Simpson, R., Kandou Labs SA (2019). Clock-embedded vector signaling codes. U.S. Patent 10,374,846. (Skliar et al., 2008, cited in "Other publications", p. 4).

[86] Shokrollahi, A., Holden, B., & Simpson, R., Kandou Labs SA (2020). Clock-embedded vector signaling codes. U.S. Patent 10,805,129. (Skliar et al., 2008, cited in "Other publications", p. 4).

[87] Shokrollahi, A., Holden, B., & Stauffer, D., Kandou Labs SA (2019). High speed memory interface. U.S. Patent 10,467,177. (Skliar et al., 2008, cited in "Other publications", p. 3).

[88] Shokrollahi, A., Hormati, A., & Tajalli, A., Kandou Labs SA (2019). Vector signaling codes for density routed wire groups. U.S. Patent 10,333,741. (Skliar et al., 2008, cited in "Other Publications", p. 6).

[89] Shokrollahi, A., Hormati, A., & Ulrich, R., Kandou Labs SA (2019). Method and apparatus for low power chip-to-chip communications with constrained ISI ratio. U.S. Patent 10,348,436. (Skliar et al., 2008, cited in "Other Publications", p. 5).

[90] Shokrollahi, A., Hormati, A., & Ulrich, R., Kandou Labs SA (2021). Method and apparatus for low power chip-to-chip communications with constrained ISI ratio. U.S. Patent 11,025,359. (Skliar et al., 2008, cited in "Other Publications", p. 3).

[91] Shokrollahi, A., Hormati, A., & Ulrich, R., Kandou Labs SA (2023). Method and apparatus for low power chip-to-chip communications with constrained ISI ratio. U.S. Patent 11,683,113. (Skliar et al., 2008, cited in "Other Publications", p. 3).

[92] Simpson, R., Kevin, A., Stewart, J., & Hormati, A., Kandou Labs SA (2017). Clock data alignment system for vector signaling code communications link. U.S. Patent 9,577,815. (Skliar et al., 2008, cited in "Other publications", p. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[93] Simpson, R., Kevin, A., Stewart, J., Holden, B., & Shokrollahi, A., Kandou Labs SA (2016). Methods and systems for energy-efficient communications interface. U.S. Patent 9,362,962. (Skliar et al., 2008, cited in "Other publications", p. 3).

[94] Simpson, R., & Ulrich, R., Kandou Labs SA (2016). Control loop management and vector signaling code communications links. U.S. Patent 9,450,744. (Skliar et al., 2008, cited in "Other publications", p. 3).

[95] Tajalli, A., Kandou Labs SA (2020). Dynamically weighted exclusive OR gate heavily weighted output segments for phase detection and phase interpolation. U.S. Patent 10,554,380. (Skliar et al., 2008, cited in "Other publications", p. 5).

[96] Tajalli, A., Kandou Labs SA (2017). Enhanced phase interpolation circuit. U.S. Patent 9,557,760. (Skliar et al., 2008, cited in "Other publications", p. 4).

[97] Tajalli, A., Kandou Labs SA (2018). High performance phase locked loop. U.S. Patent 10,057,049. (Skliar et al., 2008, cited in "Other publications", p. 5).

[98] Tajalli, A., Kandou Labs SA (2018). Lock detector for phase lock loop. U.S. Patent 9,906,358. (Skliar et al., 2008, cited in "Other publications", p. 5).

[99] Tajalli, A., Kandou Labs SA (2019). Lock detector for phase lock loop. U.S. Patent 10,355,852. (Skliar et al., 2008, cited in "Other publications", p. 5).

[100] Tajalli, A., Kandou Labs SA (2019). Methods and systems for providing multi-stage distributed decision feedback equalization. U.S. Patent 10,326, 623. (Skliar et al., 2008, cited in "Other publications", p. 5).

[101] Tajalli, A., Kandou Labs SA (2019). Multiphase data receiver with distributed DFE. U.S. Patent 10,372,665. (Skliar et al., 2008, cited in "Other publications", p. 5).

[102] Tajalli, A., Kandou Labs SA (2019). Multistage sampler with increased gain. U.S. Patent 10,200,218. (Skliar et al., 2008, cited in "Other publications", p. 5).

[103] Tajalli, A., Kandou Labs SA (2019). Phase rotation circuit for eye scope measurements. U.S. Patent 10,277,431 (Skliar et al., 2008, cited in "Other publications", p. 5).

[104] Tajalli, A., Kandou Labs SA (2018). Sampler with low input kickback. U.S. Patent 10,003,454. (Skliar et al., 2008, cited in "Other publications", p. 5).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[105] Tajalli, A., Kandou Labs SA (2016). Symmetric linear equalization circuit with increased gain. U.S. Patent 9,419,564. (Skliar et al., 2008, cited in "Other publications", p. 3).

[106] Tajalli, A., Kandou Labs SA (2017). Symmetric linear equalization circuit with increased gain. U.S. Patent 9,692,381. (Skliar et al., 2008, cited in "Other publications", p. 4).

[107] Tajalli, A., Kandou Labs SA (2020). Synchronously-switched multi-input demodulating comparator. U.S. Patent 10,693,688. (Skliar et al., 2008, cited in "Other publications", p. 3).

[108] Tajalli, A., Kandou Labs SA (2021). Synchronously-switched multi-input demodulating comparator. U.S. Patent 11,063,799. (Skliar et al., 2008, cited in "Other publications", p. 4).

[109] Tajalli, A., Kandou Labs SA (2022). Synchronously-switched multi-input demodulating comparator. U.S. Patent 11,469,931. (Skliar et al., 2008, cited in "Other publications", p. 4).

[110] Tajalli, A., Kandou Labs SA (2018). Voltage sampler driver with enhanced high-frequency gain. U.S. Patent 10,oo3,315. (Skliar et al., 2008, cited in "Other publications", p. 5).

[111] Tajali, A., Conrie, H., & Shokrollahi, A., Kandou Labs SA (2016). Efficient processing and detection of balanced codes. U.S. Patent 9,413,384. (Skliar et al., 2008, cited in "Other publications", p. 3).

[112] Tajali, A., Conrie, H., & Shokrollahi, A., Kandou Labs SA (2017). Efficient processing and detection of balanced codes. U.S. Patent 9,577,664. (Skliar et al., 2008, cited in "Other publications", p. 4).

[113] Tajalli, A., & Hormati, A., Kandou Labs SA (2020). Multi-modal data-driven clock recovery circuit. U.S. Patent 10,693,473. (Skliar et al., 2008, cited in "Other publications", p. 4).

[114] Tajalli, A., & Hormati, A., Kandou Labs SA (2022). Multi-modal data-driven clock recovery circuit. U.S. Patent 11,271,571. (Skliar et al., 2008, cited in "Other publications", p. 4).

[115] Tajalli, A., & Hormati, A., Kandou Labs SA (2023). Multi-modal data-driven clock recovery circuit. U.S. Patent 11,804,845. (Skliar et al., 2008, cited in "Other publications", p. 5).

[116] Tajalli, A., & Shokrollahi, A., Kandou Labs SA (2019). Quadrature duty cycle error in matrix phase lock lock loop. U.S. Patent 10,200,288. (Skliar et al., 2008, cited in "Other publications", p. 5).

[117] Ulrich, R., Kandou Labs SA (2018). Data-driven voltage regulator. U.S. Patent 9,985,634. (Skliar et al., 2008, cited in "Other publications", p. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[118] Ulrich, R., Kandou Labs SA (2017). Multilevel driver for high speed chip-to-chip communications. U.S. Patent 9,544,015. (Skliar et al., 2008, cited in "Other publications", p. 4).

[119] Ulrich, R., Kandou Labs SA (2018). Multilevel driver for high speed chip-to-chip communications. U.S. Patent 9,917,711. (Skliar et al., 2008, cited in "Other publications", p. 5).

[120] Ulrich, R., Kandou Labs SA (2018). Multilevel driver for high speed chip-to-chip communications. U.S. Patent 10,091,033. (Skliar et al., 2008, cited in "Other publications", p. 5).

[121] Ulrich, R., & Hunt, P., Kandou Labs SA (2016). Circuits for efficient detection of vector signaling codes for chip-to-chip communication. U.S. Patent 9,450,791. (Skliar et al., 2008, cited in "Other publications", p. 3).

[122] Ulrich, R., & Hunt, P., Kandou Labs SA (2017). Circuits for efficient detection of vector signaling codes for chip-to-chip communication. U.S. Patent 9,819,522. (Skliar et al., 2008, cited in "Other publications", p. 5).

[123] Ulrich, R., & Hunt, P., Kandou Labs SA (2018). Circuits for efficient detection of vector signaling codes for chip-to-chip communication. U.S. Patent 10,164,809. (Skliar et al., 2008, cited in "Other publications", p. 6).

[124] Ulrich, R., & Hunt, P., Kandou Labs SA (2016). Circuits for efficient detection of vector signaling codes for chip-to-chip communication using sums of differences. U.S. Patent 9,288,082. (Skliar et al., 2008, cited in "Other publications", p. 3).

[125] Ulrich, R., & Shokrollahi, A., Kandou Labs SA (2016). Bus reversible orthogonal differential vector signaling codes. U.S. Patent 9,432,082. (Skliar et al., 2008, cited in "Other publications", p. 3).

[126] Ulrich, R., & Shokrollahi, A., Kandou Labs SA (2018). Bus reversible orthogonal differential vector signaling codes. U.S. Patent 10,003,424. (Skliar et al., 2008, cited in "Other publications", p. 5).

[127] Ulrich, R., & Shokrollahi, A., Kandou Labs SA (2019). Bus reversible orthogonal differential vector signaling codes. U.S. Patent 10,404,394. (Skliar et al., 2008, cited in "Other publications", p. 4).

[128] Ulrich, R., Tajalli, A. Hormati, A. & Simpson, R., Kandou Labs SA (2020). Method for measuring and connecting multi-wire screw. U.S. Patent 10,686,583. (Skliar et al., 2008, cited in "Other publications", p. 5).

[129] Walter, C., Kandou Labs SA (2014). Multiwire linear equalizer for vector signaling code receiver. U.S. Patent 9,419,828. (Skliar et al., 2008, cited in "Other publications", p. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.13.566944; this version posted November 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[130] Wiebe, N., & Roetteler, M., Microsoft Technology Licensing, LLC (2019). Quantum algorithms for arithmetic and function synthesis, U.S. Patent 10,320,360. (Skliar et al., 2008, cited in "Other publications", p. 2).

[131] Nurk, S. et al. (2022). The complete sequence of a human genome. *Science*, Vol. 376, pp. 44-53. DOI:10.1126/science.abj6987.

[132] T2T-CHM13v2.0 – Genome – Assembly – NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCF_009914755.1/.

[133] Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S. K., Tourna, A., Yakovleva, A., Palmieri, T. & Ciccarelli, F. D. (2019). The Network of Cancer Genes (NCG): A comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biology*, Vol. 20(1). DOI: https://doi.org/10.1186/s13059-018-1612-0.