



- (51) **International Patent Classification:**  
*G06F 19/24* (2011.01) *G06F 19/20* (2011.01)
- (21) **International Application Number:**  
PCT/US2016/035870
- (22) **International Filing Date:**  
3 June 2016 (03.06.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
62/171,855 5 June 2015 (05.06.2015) US
- (71) **Applicant:** LIFE TECHNOLOGIES CORPORATION  
[US/US]; 5823 Newton Drive, Carlsbad, California 92008 (US).
- (72) **Inventors:** MAJUMDAR, Nivedita Sumi; Life Technologies Corporation, 5823 Newton Drive, Carlsbad, California 92008 (US). WESSEL, Thomas; Life Technologies Corporation, 5823 Newton Drive, Carlsbad, California 92008 (US).

- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

- (54) **Title:** DETERMINING THE LIMIT OF DETECTION OF RARE TARGETS USING DIGITAL PCR

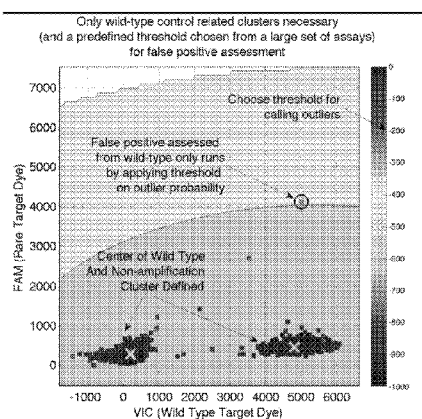


FIG. 1B

(57) **Abstract:** A method for determining false positives calls in a biological data plot is provided. The method includes identifying a first data cluster as non-amplification data points within the biological data plot and identifying a second data cluster as wild-type positives within the biological data plot. The method further includes estimating a position in the biological data plot of a center of the first and second data clusters. The method further includes determining, for each data point within the first data cluster, a probability of belonging to the first data cluster and determining, for each data point within the second data cluster, a probability of belonging to the second data cluster. The method includes applying a probability threshold for each data point within the first and second data cluster to identify false positives.



---

**Declarations under Rule 4.17:**

— *of inventorship (Rule 4.17(iv))*

**Published:**

— *with international search report (Art. 21(3))*

## DETERMINING THE LIMIT OF DETECTION OF RARE TARGETS USING DIGITAL PCR

## BACKGROUND

**[0001]** Detection and quantification of mutant alleles in tumor tissue is important to cancer research. Testing for the presence of mutations in circulating free DNA (cfDNA) is one of the less invasive research methods available at this time. Digital PCR presents a research tool for mutation detection in cfDNA at a sensitivity level of 1% and below.

**[0002]** The digital method segments sample DNA into a large number of reaction partitions. Upon performing PCR, amplification is detected in reactions with DNA template and no amplification is detected in reactions lacking DNA template. This large scale partitioning isolates the rare target within a subset of partitions, elevates the rare to wild-type ratio within any specific partition (compared to the original PCR mix), and enhances the amplification probability and detectability of the rare target. These three effects enable detection of the rare target with high sensitivity.

**[0003]** Challenges associated with digital PCR experiments for rare allele detection include understanding the limit of detection of the assay and platform. Data points corresponding to rare target are by definition far fewer than the data points corresponding to positives for the wild-type target. This makes identification of the rare target challenging. One known approach to addressing this challenge requires overlaying wild-type control data with positive control data to guide the definition for a boundary of the wild-type event in fluorescence space. The data points outside of this boundary are considered true positives for the rare target for unknown sample (and false positives for a control sample with wild-type only target). This strategy only works when the inter-run variation in signal levels is negligible and/or when a specific normalization is applied to account for such variation.

**[0004]** A need therefore exists for a more robust approach that works independent of interrun variations in signal levels.

## SUMMARY

**[0005]** In one exemplary embodiment, a method for determining false positives calls in a biological data plot is provided. The method includes identifying a first data cluster as non-amplification data points within the biological data plot and identifying a second data cluster as wild-type positives within the biological data plot. The method further includes estimating a position in the biological data plot of a center of the first and second data clusters. The method further includes determining, for each data point within the first data cluster, a probability of belonging to the first data cluster and determining, for each data point within the second data cluster, a probability of belonging to the second data cluster. The method includes applying a probability threshold for each data point within the first and second data cluster to identify false positives.

## DESCRIPTION OF THE FIGURES

**[0006]** FIG. 1A illustrates a known method (Method A) identifying false positive events when in the detection of rare targets according to various embodiments described herein.

**[0007]** FIG. 1B illustrates a method of the present disclosure (Method B) for identifying false positive events when in the detection of rare targets according to various embodiments described herein.

**[0008]** FIGS. 2A-2B illustrate duplicate runs at 0% target to total ratio (control) according to various embodiments described herein.

**[0009]** FIGS. 3A-3B illustrate duplicate runs at 10% target to total ratio according to various embodiments described herein.

**[0010]** FIGS. 4A-4B illustrate duplicate runs at 1% target to total ratio according to various embodiments described herein.

**[0011]** FIGS. 5A-5B illustrate duplicate runs at 0.1% target to total ratio according to various embodiments described herein.

**[0001]** FIG. 6 illustrates an exemplary computing system for implementing various embodiments described herein.

**[0002]** FIG. 7 illustrates an exemplary distributed network system according to various embodiments described herein.

## DETAILED DESCRIPTION

[0012] Data points corresponding to rare target are by definition far fewer than the data points corresponding to positives for the wild-type target. This makes identification of the rare target challenging. A known method (method A) for solving this challenge is described below.

**Known Method A**

[0013] The data from the wild-type control is overlaid with the data from the positive control to guide the definition for a boundary of the wild-type event in fluorescence space. The data points outside of this boundary are considered true positives for the rare target for unknown sample (and false positives for a control sample with wild-type only target). This strategy works when the inter-run variation in signal levels is negligible or when a specific normalization is applied to account for such variation.

[0014] A more reliable method (Method B), according to various embodiments, for solving this challenge is described below:

**Method B**

[0015] A second approach, described in the present disclosure, identifies the center of the non-amplification cluster and of the wild-type positive cluster. This approach next evaluates, for each data point, the probabilities  $\{p_1, p_2\}$  of belonging to either of these clusters. The final step establishes, again for each data point, a single probability,  $p = \max\{p_1, p_2\}$ , upon which a threshold may be applied to identify outlier events that do not belong within one of these main clusters. This strategy is more robust as it works independent of inter-run variations in signal levels. It is based on the assumption of finding a sizable non-amplification and wildtype positive clusters.

[0016] If false positives are identified using control chips, lower limits on detectable concentration of the rare target can be established. Replicate runs may be used to get an understanding of the distribution of false positive events for a given assay system. Then, a lower limit of detection (above the false positive rate) of the assay system can be calculated.

[0017] The present disclosure relates, in some embodiments, to a method for identifying false positive events in the detection of rare targets. The method can include identifying the center of the non-amplification cluster and of the wild-type positive cluster. This method can also include evaluating, for each data point, the probabilities (e.g.,  $\{p_1, p_2\}$ ) of belonging to either of the

identified clusters. This method can further include establishing, for each data point, a single probability (e.g.,  $p = \max\{p1, p2\}$ ) upon which a threshold may be applied to identify outlier events that do not belong within one of these main clusters.

**[0018]** This method is more robust than known approaches to false positive assessment strategies as it works independent of interrun variations in signal levels. It is based, among other things, on an assumption of finding a sizable non-amplification and wild-type positive clusters.

**[0019]** If false positives are identified using control chips, lower limits on detectable concentration of the rare target can be established. Replicate runs may be used to get an understanding of the distribution of false positive events for a given assay system. Then, a lower limit of detection (above the false positive rate) of the assay system can be calculated.

**[0020]** This description below compares a known false positive assessment method to a method of the present disclosure, using the signal levels of the no-amplification cluster and the wild-type cluster where available. Once the false positive call rate is established, this description below outlines a method to determine the limit of detection of the assay and platform, at a given level of confidence. Given the number of partitions, the interrogated volume and the false call rate, the tradeoffs between sample load and sensitivity are also discussed.

**[0021]** The mathematics outlined to calculate the theoretical limit of detection is applied on a set of assays covering the KRAS codon mutations commonly found in tumor tissues. Experimental results showing a detection of at least 0.1% mutation rate are presented as examples. Test samples were created using both mutant plasmid and mutant genomic DNA mixed with wild-type genomic DNA at a predefined percentage.

### **Experimental Design Considerations**

**[0022]** While the false positive rate puts a lower limit on the concentration of rare targets that can be reliably measured, there are two other considerations for sensitivity: 1) The larger the interrogated volume, the higher the sensitivity (or the lower the concentration that you can detect); and 2) The minimum in-partition rare to wild-type ratio that can be tolerated by the assay dictates how much wild-type target may be loaded on to the chip.

## Experimental Protocol

**[0023]** Materials: 0.1x TE Buffer from 1x TE Buffer, 6.8 ng/uL gDNA from 100 ng/uL or 10 ng/uL gDNA, “1X” plasmid from “10X” plasmid, sample plate (e.g., QuantStudio™ 3D Chips), a PCR thermal cycler, and a fluorescence detection or reader instrument (e.g., QuantStudio™ 3D instrument).

**[0024]** Mixture Creation: Prep loading mixture for “10%” chips: In a labeled tube (1.5 mL or 0.5 mL), pipet in the following: 40 µL of Master Mix, 20 µL of 6.8 ng/µl gDNA, 16 uL of “10X” plasmid, 4 µL of the 20X rare mutation assay. Vortex the finished tube. For 1% chips, dilute the plasmid to a “1X” tube and use 16 µL of the “1X”. For wild type chips, replace the 16 µL of plasmid with 16 µL of ultrapure water.

**[0025]** Run: Load 14.5 µL on each sample plate (e.g., QuantStudio™ 3D chip) and thermal cycle per the rare mutation assay thermal cycling conditions prior to imaging on the fluorescence detection or reader instrument (e.g., QuantStudio™ 3D instrument), following the protocol prescribed for rare mutation assays.

## Analysis Protocol: False Positive Identification

**[0026]** FIG. 1A and FIG. 1B shows two methods for identifying false positives from non-template controls and wild-type control runs. FIG. 1A illustrates a known method (Method A) of designating the non-amplification and wild-type positive cluster area in fluorescence space by explicit boundary, and designating points outside of this area designated as false positives. FIG. 1B illustrates a method of the present disclosure for identifying false positives, the method including estimating cluster centers and spread respectively from the non-amplification and wild-type positives, fitting to a two dimensional Gaussian model, and applying a threshold on log probability for belonging to modeled cluster to identify false positives.

**[0027]** As discussed above, it is a challenge to draw boundaries where the density of points is low, trying to decide whether or not a point on the edge of a cluster is a real positive or not, as necessary to apply Method A. This method only works when the inter-run variation in signal levels is negligible and/or when a specific normalization is applied to account for such variation. On the other hand, Method B can require identification of centers of clusters that have significant membership.

**[0028]** Equation set 1 below describes the model used to calculate the likelihood of outlier status for a given data point, when both the non-amplification cluster and the wild-type positive cluster exists (wild-type control). This can easily be generalized to the case where only the non-amplification cluster exists (non-template control).

**[0029]** Let the probabilities  $p_1$  and  $p_2$  denote the probability of belonging with the non-amplification and the wild-type positive cluster respectively.

$$p_1(v, f) = C \times \exp \left[ -\frac{1}{2} A \Sigma_A^{-1} A^T \right]$$

$$p_2(v, f) = C \times \exp \left[ -\frac{1}{2} B \Sigma_B^{-1} B^T \right]$$

where:

$C$  is the constant associated with the 2D Gaussian modeling (Here,  $C=1$ )

$A = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the non-amplification cluster 1

$B = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the wild-type positive cluster

$\Sigma$  is the covariance matrix  $\begin{pmatrix} \text{var}(f) & \text{cov}(f, v) \\ \text{cov}(f, v) & \text{var}(v) \end{pmatrix}$  with  $\Sigma_A$  calculated from the non-amplification cluster and  $\Sigma_B$  calculated from the wild-type positive cluster respectively.

$$p(v, f) = \max(p_1, p_2)$$

**[0030]** A set of 42 TaqMan® assays were chosen with 4 replicate runs of the wild-type control. Positive controls at 1 to 10% titration of the mutant alleles to fixed concentration of the wild-type allele were also run for these assays. Based upon this data, a threshold of -200 on  $\log(p)$  is chosen to identify a true false positive distinct from the scatter at the periphery of the wildtype cluster. A true false positive is a positive on a control that would cluster with true rare target positives).

**[0031]** Apart from signal strength (method A), and separation from main clusters (method B), one last factor to consider for false positive determination is the through-hole level quality value of the specific point and its neighboring points, if working with an array based technology where

this information is available, such as the QuantStudio 3D platform. Using high quality data points (or points from a high data quality region) is recommended.

### Analysis Protocol: Estimating the False Positive Rate and the Limit of Detection

[0032] Once the number of false positives for the  $i^{\text{th}}$  run is available, it is normalized by the wild-type load per equation 2 (from Coren A. Milbury, Qun Zhong, Jesse Lin, Miguel Williams, Jeff Olson, Darren R. Link, Brian Hutchison. “Determining the lower limits of detection of digital PCR assays for cancer-related gene mutations.” *Biomolecular Detection and Quantification*. Volume 1, Issue 1. September 2014, Pages 8 – 22) (hereinafter “*Milbury*”). 2

$$[0033] \text{Normalized \#False Positive}_i = \left( \frac{1}{k} \sum_{\text{run\# } j=1}^k \frac{\gamma_{\text{mutant}}^j}{\gamma_{\text{wild-type}}^j} \right) \times \gamma_{\text{wild-type}}^i \times N_i$$

[0034] And then the lowest limit of detection for that assay system is determined per equation set 3 (see, *Milbury*).

| $\Lambda_{FP}$ | $LoB$                                           | $LoD$                                   |
|----------------|-------------------------------------------------|-----------------------------------------|
| 0              | 0                                               | 3                                       |
| 0 – 0.5        | 1                                               | 5                                       |
| > 0.05         | $\Lambda_{FP} + 1.645 \sqrt{\Lambda_{FP}} + .8$ | $(1.645 + \sqrt{1.645^2 + 4LoB^2}) / 4$ |

3

[0035] where  $\Lambda_{FP}$  is the normalized average number of false positives per run,  $LoB$  is the limit of blank and  $LoD$  is the limit of detection.

[0036] Note that knowing the average number of false positives does not allow us to correct an answer when evaluating unknown targets. At a given run, the actual number of false positives can take any value. Therefore the best use of the false positive rate knowledge is for determining what the minimum number of events above which we can reliably conclude that the observed set of data is different from the false positive distribution.

### Results

[0037] Ten runs from KRAS 516 are annotated by manual calling. Method B is then deployed to estimate the rare dye. The estimated number is compared to the annotation result and shows good correspondence (Table 1).

**Table 1: Results using a candidate assay design targeting KRAS 516**

| Chip # | Task                                            | Wild Type Copies/ $\mu$ L | # Mutant (annotated) | Mutant Copies/ $\mu$ L (annotated) | # Mutant (Method B) | Mutant Copies/ $\mu$ L (Method B) | Normalized Number of False Positive |
|--------|-------------------------------------------------|---------------------------|----------------------|------------------------------------|---------------------|-----------------------------------|-------------------------------------|
| 1      | Unknown                                         | 51.75                     | 325                  | 20.47                              | 324                 | 20.41                             |                                     |
| 2      | Unknown                                         | 64.11                     | 308                  | 20.48                              | 295                 | 19.68                             |                                     |
| 3      | Unknown                                         | 65.15                     | 333                  | 22.92                              | 331                 | 22.79                             |                                     |
| 4      | Unknown                                         | 61.11                     | 30                   | 1.98                               | 31                  | 2.04                              |                                     |
| 5      | Unknown                                         | 54.67                     | 39                   | 2.69                               | 41                  | 2.83                              |                                     |
| 6      | Unknown                                         | 59.85                     | 34                   | 2.28                               | 34                  | 2.28                              |                                     |
| 7      | Wild-type                                       | 50.81                     | 1                    | 0.06                               | 10                  | 0                                 | 1.54                                |
| 8      | Wild-type                                       | 59.54                     | 2                    | 0.16                               | 1                   | 0.08                              | 1.45                                |
| 9      | Wild-type                                       | 51.05                     | 1                    | 0.07                               | 1                   | 0.07                              | 1.50                                |
| 10     | Wild-type                                       | 58.83                     | 2                    | 0.15                               | 2                   | 0.15                              | 1.52                                |
|        | Average False Positive Rate from Wild-type Runs |                           |                      |                                    |                     |                                   | 1.51                                |
|        | Lowest Limit of Detection at 95% Confidence     |                           |                      |                                    |                     |                                   | 3.85                                |

**[0038]** Wild-type only control, and rare mutation at set proportions to the wild type were run for assays targeting the KRAS 521 in duplicates. In particular, similar experiments were run for KRAS 521 in duplicates at 0% (Wild-type only control), 0.1%, 1%, and 10% target to total ratios. Provided in FIGS. 2-5 are data from each run and the quantifications based upon manual calling compared to method B. Again, we see good agreement. In particular, FIGS. 2A-2B illustrate duplicate runs at 0% target to total ratio (control), FIGS. 3A-3B illustrate duplicate runs at 10% target to total ratio, FIGS. 4A-4B illustrate duplicate runs at 1% target to total ratio, and FIGS. 5A-5B illustrate duplicate runs at 0.1% target to total ratio. In each FIG., rare target

quantification by manual setting of threshold, indicated by the lines (Method A), match well with those predicted by Method B (indicated by \* symbol) yielding up to 0.1% rare mutation detection.

**[0039]** As detailed in the foregoing, one can evaluate a signal level above which a data point will be considered as a positive, typically done using both positive and wild-type controls as described in Method A. This is susceptible to run to run variation in signal levels. This disclosure introduces an alternate method based upon the assumption that there is sufficient numbers of points belonging to the non-amplification cluster and the positive cluster for the wild-type target (unless the run is a no template control, in which case you only have the non-amplification cluster). The statistics of these one or two dominant clusters are used to assess if a given point belongs with these cluster or not. If not, they are suitable to be labeled as outliers or false positives, as described by Method B. The efficacy of this method is demonstrated by the predicting of the rare concentration correctly where the true rare data points have been manually annotated. Once the number of false positives are determined, they are normalized across replicates by known methods (see, for example, those recommended in Milbury), and based upon the normalized rate, the lowest limit of detection is also evaluated as described in, for example, Milbury.

**[0040]** Those skilled in the art will recognize that the operations of the various embodiments may be implemented using hardware, software, firmware, or combinations thereof, as appropriate. For example, some processes can be carried out using processors or other digital circuitry under the control of software, firmware, or hard-wired logic. (The term “logic” herein refers to fixed hardware, programmable logic and/or an appropriate combination thereof, as would be recognized by one skilled in the art to carry out the recited functions.) Software and firmware can be stored on non-transitory computer-readable media. Some other processes can be implemented using analog circuitry, as is well known to one of ordinary skill in the art. Additionally, memory or other storage, as well as communication components, may be employed in embodiments of the present teachings.

**[0041]** FIG. 6 is a block diagram that illustrates a computer system 600 that can be employed to carry out processing functionality, and to implement various components or subsystems of the systems described herein according to various embodiments. For example, system 600 can

comprise all or apportion of devices 540, client devices, 502, 512, or 530, servers 522, etc. Computing system 600 can include one or more processors, such as a processor 604. Processor 604 can be implemented using a general or special purpose processing engine such as, for example, a microprocessor, controller or other control logic. In this example, processor 604 is connected to a bus 602 or other communication medium.

**[0042]** Further, it should be appreciated that a computing system 600 of FIG. 6 can be embodied in any of a number of forms, such as a rack-mounted computer, mainframe, supercomputer, server, client, a desktop computer, a laptop computer, a tablet computer, hand-held computing device (e.g., PDA, cell phone, smart phone, palmtop, etc.), cluster grid, netbook, embedded systems, or any other type of special or general purpose computing device as may be desirable or appropriate for a given application or environment. Additionally, a computing system 600 can include a conventional network system including a client/server environment and one or more database servers, or integration with LIS/LIMS infrastructure. A number of conventional network systems, including a local area network (LAN) or a wide area network (WAN), and including wireless and/or wired components, are known in the art. Additionally, client/server environments, database servers, and networks are well documented in the art. According to various embodiments described herein, computing system 600 may be configured to connect to one or more servers in a distributed network. Computing system 600 may receive information or updates from the distributed network. Computing system 600 may also transmit information to be stored within the distributed network that may be accessed by other clients connected to the distributed network.

**[0043]** Computing system 600 may include bus 602 or other communication mechanism for communicating information, and processor 604 coupled with bus 602 for processing information.

**[0044]** Computing system 600 also includes a memory 606, which can be a random access memory (RAM) or other dynamic memory, coupled to bus 602 for storing instructions to be executed by processor 604. Memory 606 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 604. Computing system 600 further includes a read only memory (ROM) 608 or other static storage device coupled to bus 602 for storing static information and instructions for processor 604.

**[0045]** Computing system 600 may also include a storage device 610, such as a magnetic disk, optical disk, or solid state drive (SSD) is provided and coupled to bus 602 for storing information and instructions. Storage device 610 may include a media drive and a removable storage interface. A media drive may include a drive or other mechanism to support fixed or removable storage media, such as a hard disk drive, a floppy disk drive, a magnetic tape drive, an optical disk drive, a CD or DVD drive (R or RW), flash drive, or other removable or fixed media drive. As these examples illustrate, the storage media may include a computer-readable storage medium having stored therein particular computer software, instructions, or data.

**[0046]** In alternative embodiments, storage device 610 may include other similar instrumentalities for allowing computer programs or other instructions or data to be loaded into computing system 600. Such instrumentalities may include, for example, a removable storage unit and an interface, such as a program cartridge and cartridge interface, a removable memory (for example, a flash memory or other removable memory module) and memory slot, and other removable storage units and interfaces that allow software and data to be transferred from the storage device 610 to computing system 600.

**[0047]** Computing system 600 can also include a communications interface 618. Communications interface 618 can be used to allow software and data to be transferred between computing system 600 and external devices. Examples of communications interface 618 can include a modem, a network interface (such as an Ethernet or other NIC card), a communications port (such as for example, a USB port, a RS-232C serial port), a PCMCIA slot and card, Bluetooth, etc. Software and data transferred via communications interface 618 are in the form of signals which can be electronic, electromagnetic, and optical or other signals capable of being received by communications interface 618. These signals may be transmitted and received by communications interface 618 via a channel such as a wireless medium, wire or cable, fiber optics, or other communications medium. Some examples of a channel include a phone line, a cellular phone link, an RF link, a network interface, a local or wide area network, and other communications channels.

**[0048]** Computing system 600 may be coupled via bus 602 to a display 612, such as a cathode ray tube (CRT) or liquid crystal display (LCD), for displaying information to a computer user.

An input device 614, including alphanumeric and other keys, is coupled to bus 602 for communicating information and command selections to processor 604, for example. An input device may also be a display, such as an LCD display, configured with touchscreen input capabilities. Another type of user input device is cursor control 616, such as a mouse, a trackball or cursor direction keys for communicating direction information and command selections to processor 604 and for controlling cursor movement on display 612. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to specify positions in a plane. A computing system 600 provides data processing and provides a level of confidence for such data. Consistent with certain implementations of embodiments of the present teachings, data processing and confidence values are provided by computing system 600 in response to processor 604 executing one or more sequences of one or more instructions contained in memory 606. Such instructions may be read into memory 606 from another computer-readable medium, such as storage device 610. Execution of the sequences of instructions contained in memory 606 causes processor 604 to perform the process states described herein. Alternatively hard-wired circuitry may be used in place of or in combination with software instructions to implement embodiments of the present teachings. Thus implementations of embodiments of the present teachings are not limited to any specific combination of hardware circuitry and software.

**[0049]** The term "computer-readable medium" and "computer program product" as used herein generally refers to any media that is involved in providing one or more sequences or one or more instructions to processor 604 for execution. Such instructions, generally referred to as "computer program code" (which may be grouped in the form of computer programs or other groupings), when executed, enable the computing system 600 to perform features or functions of embodiments of the present teachings described herein. These and other forms of non-transitory computer-readable media may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, solid state, optical or magnetic disks, such as storage device 610. Volatile media includes dynamic memory, such as memory 606. Transmission media includes coaxial cables, copper wire, and fiber optics, including the wires that comprise bus 602.

**[0050]** Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, PROM, and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave as described hereinafter, or any other medium from which a computer can read.

**[0051]** Various forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to processor 604 for execution. For example, the instructions may initially be carried on magnetic disk of a remote computer. The remote computer can load the instructions into its dynamic memory and send the instructions over a telephone line using a modem. A modem local to computing system 600 can receive the data on the telephone line and use an infra-red transmitter to convert the data to an infra-red signal. An infra-red detector coupled to bus 602 can receive the data carried in the infra-red signal and place the data on bus 602. Bus 602 carries the data to memory 606, from which processor 604 retrieves and executes the instructions. The instructions received by memory 606 may optionally be stored on storage device 610 either before or after execution by processor 604.

**[0052]** It will be appreciated that, for clarity purposes, the above description has described embodiments with reference to different functional units and processors. However, it will be apparent that any suitable distribution of functionality between different functional units, processors or domains may be used without detracting from the embodiments of the present teachings. For example, functionality illustrated to be performed by separate processors or controllers may be performed by the same processor or controller. Hence, references to specific functional units are only to be seen as references to suitable means for providing the described functionality, rather than indicative of a strict logical or physical structure or organization.

**[0053]** FIG. 7 is a diagram illustrating an example system 700 configured in accordance with one example embodiment. In system 700, one or more servers 722 can be configured to run the analysis applications for analyzing data sets produced by one or more devices or modalities 740. The data included in the data sets can be stored in one or more storage devices 750. Once the data sets have been uploaded to servers 722, then a plurality of applications running on servers 722 can be used to manipulate, analyze and visualize the data sets from anywhere. For example,

local client devices 730 can be used to access servers 722, e.g., through a hub or router 726. At the same time, the data can be accessed remotely through remote clients devices 702, which are interfaced with servers 722, e.g., via a gateway/hub/tunnel-server/etc. 710, which is itself connected to the internet 708 via some internet service provider (ISP) connection 710, or remote client servers 712, which are interfaced with servers 722, e.g., via the internet 708 and via an ISP connection 714.

**[0054]** It should also be noted that devices 740 can be directly interfaced with servers 722, e.g., through the internet. In such embodiments, the collection application and functionality can reside on servers 722, on devices 740, or both. In other embodiments, devices 740 can be interfaced with client devices 702 or 712. In such embodiments, the collection application or functionality can be included on client devices 702 or 712, devices 740, or both.

**[0055]** Client devices 702, 712, and 730 can be any kind of computing device that can be used to access servers 722. As such, these devices can be laptop, desktop, or palmtop computers, terminals, mobile computing devices such as smartphones or tablets, etc. Servers 722 can comprise one or more processors, servers, routers, co-processors, user interfaces, etc., whether co-located or located in different locations. In short, servers 722 can comprise all of the resources, both hardware and software, needed to perform the functions described herein. A more detailed description of a computer system and the resources that can be used to implement the components illustrated in FIG. 7 is described below with respect to FIG. 6.

**[0056]** Although various embodiments have been described with respect to certain exemplary embodiments, examples, and applications, it will be apparent to those skilled in the art that various modifications and changes may be made without departing from the present teachings.

## CLAIMS

What is claimed is:

1. A method for determining false positives calls in a biological data plot, the method comprising:
  - identifying a first data cluster as non-amplification data points within the biological data plot;
  - identifying a second data cluster as wild-type positives within the biological data plot;
  - estimating a position in the biological data plot of a center of the first and second data clusters;
  - determining, for each data point within the first data cluster, a probability of belonging to the first data cluster;
  - determining, for each data point within the second data cluster, a probability of belonging to the second data cluster; and
  - applying a probability threshold for each data point within the first and second data cluster to identify false positives.
2. The method of claim 1, wherein the probability of belonging to the first data cluster and the probability of belonging to the second data cluster is:

$$p_1(v, f) = C \times \exp\left[-\frac{1}{2} A \Sigma_A^{-1} A^T\right]$$

$$p_2(v, f) = C \times \exp\left[-\frac{1}{2} B \Sigma_B^{-1} B^T\right]$$

where:

$p_1$  is the probability of belonging to the first data cluster and  $p_2$  is the probability of belonging to the second data cluster,

$C$  is the constant associated with a 2D Gaussian modeling

$A = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the non-amplification cluster

$B = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the wild-type positive cluster

$\Sigma$  is the covariance matrix  $\begin{pmatrix} \text{var}(f) & \text{cov}(f,v) \\ \text{cov}(f,v) & \text{var}(v) \end{pmatrix}$  with  $\Sigma_A$  calculated from the non-amplification cluster and  $\Sigma_B$  calculated from the wild-type positive cluster respectively.

$$p(v, f) = \max(p1, p2)$$

3. The method of claim 2, wherein  $C=1$ .

4. The method of claim 1, further comprising:

determining a measure of spread between the estimated positions of the centers of the first and second data clusters.

5. A computer-readable storage medium comprising instructions, executable by a processor, for determining false positives calls in a biological data plot, the instructions comprising instructions for:

identifying a first data cluster as non-amplification data points within the biological data plot;

identifying a second data cluster as wild-type positives within the biological data plot;

estimating a position in the biological data plot of a center of the first and second data clusters;

determining, for each data point within the first data cluster, a probability of belonging to the first data cluster;

determining, for each data point within the second data cluster, a probability of belonging to the second data cluster; and

applying a probability threshold for each data point within the first and second data cluster to identify false positives.

6. The computer-readable storage medium of claim 5, wherein the probability of belonging to the first data cluster and the probability of belonging to the second data cluster is:

$$p_1(v, f) = C \times \exp \left[ -\frac{1}{2} A \Sigma_A^{-1} A^T \right]$$

$$p_2(v, f) = C \times \exp \left[ -\frac{1}{2} B \Sigma_B^{-1} B^T \right]$$

where:

p1 is the probability of belonging to the first data cluster and p2 is the probability of belonging to the second data cluster,

C is the constant associated with a 2D Gaussian modeling

$A = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the non-amplification cluster 1

$B = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the wild-type positive cluster

$\Sigma$  is the covariance matrix  $\begin{pmatrix} \text{var}(f) & \text{cov}(f,v) \\ \text{cov}(f,v) & \text{var}(v) \end{pmatrix}$  with  $\Sigma_A$  calculated from the non-amplification cluster and  $\Sigma_B$  calculated from the wild-type positive cluster respectively.

$$p(v, f) = \max(p1, p2)$$

7. The computer-readable storage medium of claim 6, wherein C=1.

8. The computer-readable storage medium of claim 5, wherein the instructions further comprise instructions for:

determining a measure of spread between the estimated positions of the centers of the first and second data clusters.

9. A system for determining false positives calls in a biological data plot, the method comprising:

a processor; and

a memory comprising instructions, executable by the processor, for:

identifying a first data cluster as non-amplification data points within the biological data plot;

identifying a second data cluster as wild-type positives within the biological data plot;

estimating a position in the biological data plot of a center of the first and second data clusters;

determining, for each data point within the first data cluster, a probability of belonging to the first data cluster;

determining, for each data point within the second data cluster, a probability of belonging to the second data cluster; and  
 applying a probability threshold for each data point within the first and second data cluster to identify false positives.

10. The system of claim 9, wherein the probability of belonging to the first data cluster and the probability of belonging to the second data cluster is:

$$p_1(v, f) = C \times \exp\left[-\frac{1}{2} A \Sigma_A^{-1} A^T\right]$$

$$p_2(v, f) = C \times \exp\left[-\frac{1}{2} B \Sigma_B^{-1} B^T\right]$$

where:

p1 is the probability of belonging to the first data cluster and p2 is the probability of belonging to the second data cluster,

C is the constant associated with a 2D Gaussian modeling

$A = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the non-amplification cluster 1

$B = \begin{pmatrix} v - \mu_v \\ f - \mu_f \end{pmatrix}$  with means calculated from the wild-type positive cluster

$\Sigma$  is the covariance matrix  $\begin{pmatrix} \text{var}(f) & \text{cov}(f, v) \\ \text{cov}(f, v) & \text{var}(v) \end{pmatrix}$  with  $\Sigma_A$  calculated from the non-amplification cluster and  $\Sigma_B$  calculated from the wild-type positive cluster respectively.

$$p(v, f) = \max(p1, p2)$$

11. The system of claim 10, wherein C=1.

12. The system of claim 9, wherein the memory further comprises instructions for:  
 determining a measure of spread between the estimated positions of the centers of the first and second data clusters.

1/12

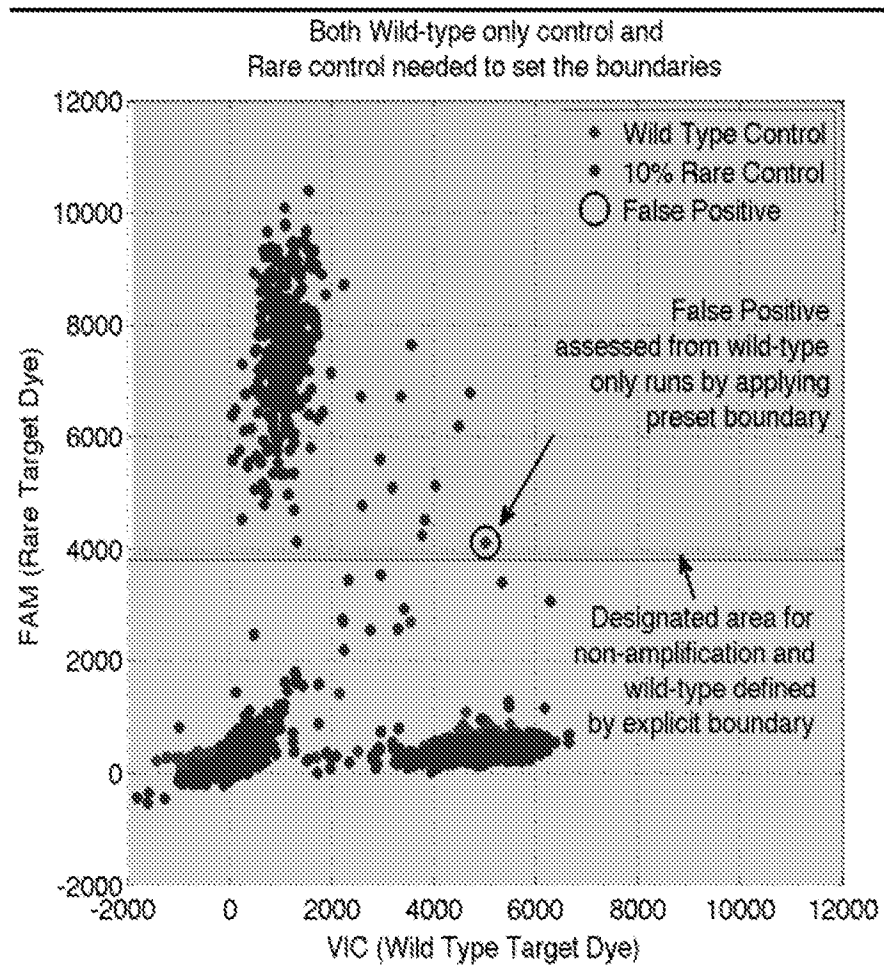


FIG. 1A

2/12

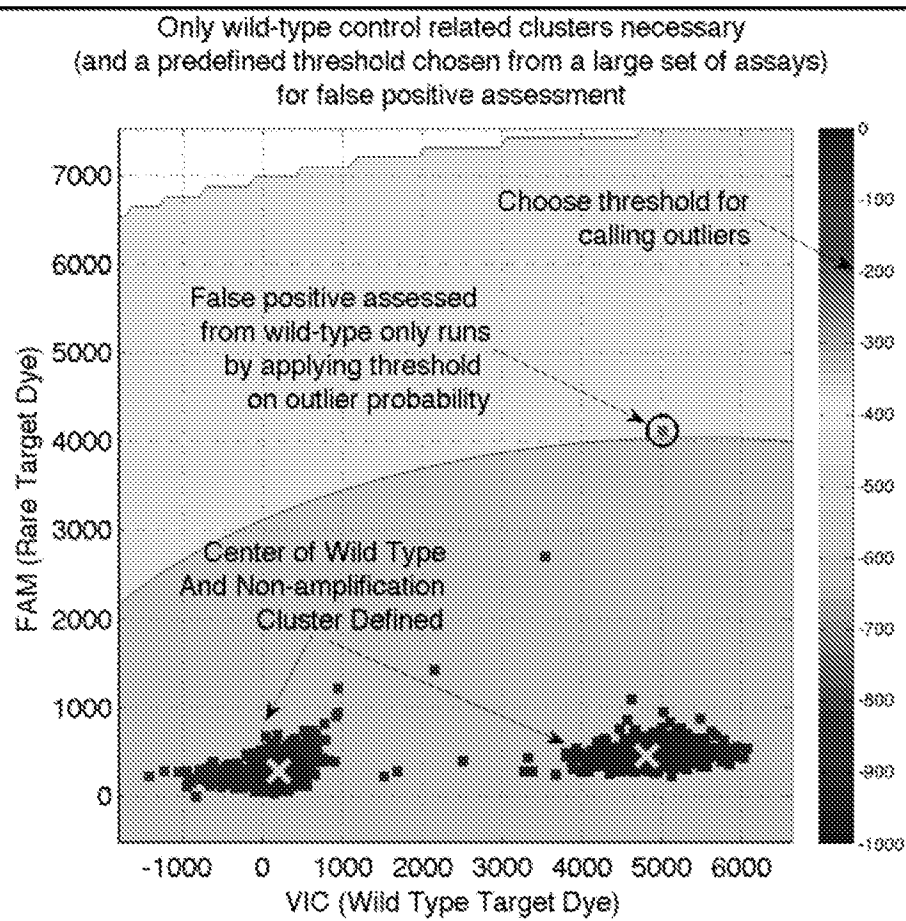


FIG. 1B

3/12

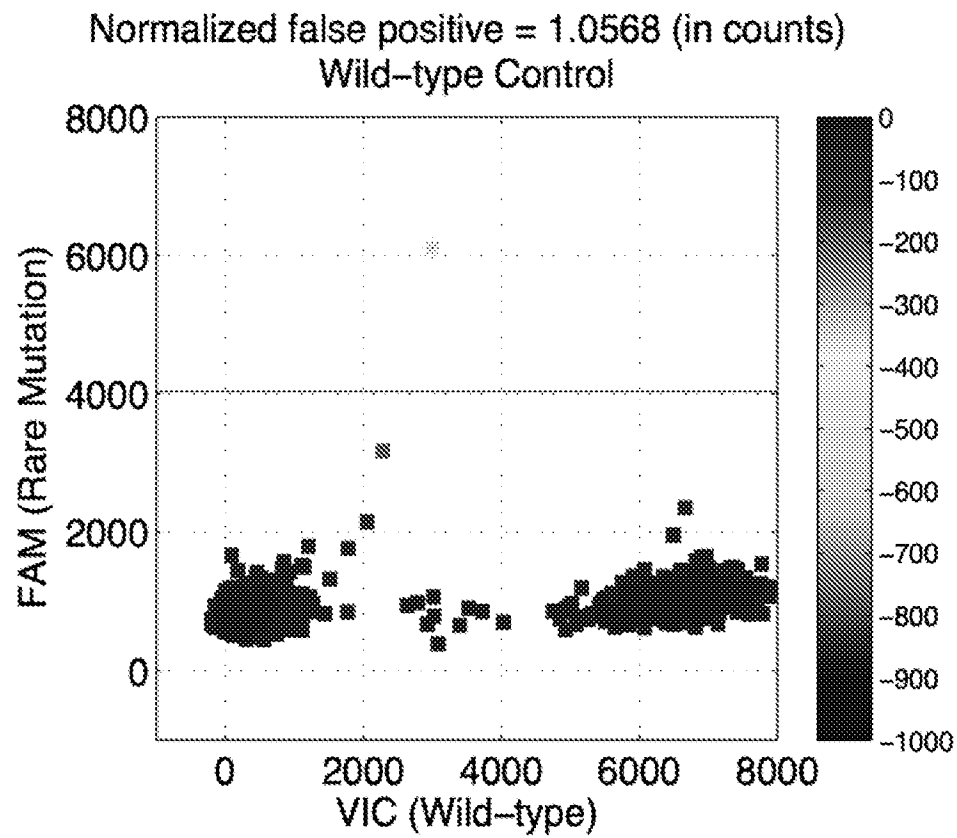


FIG. 2A

4/12

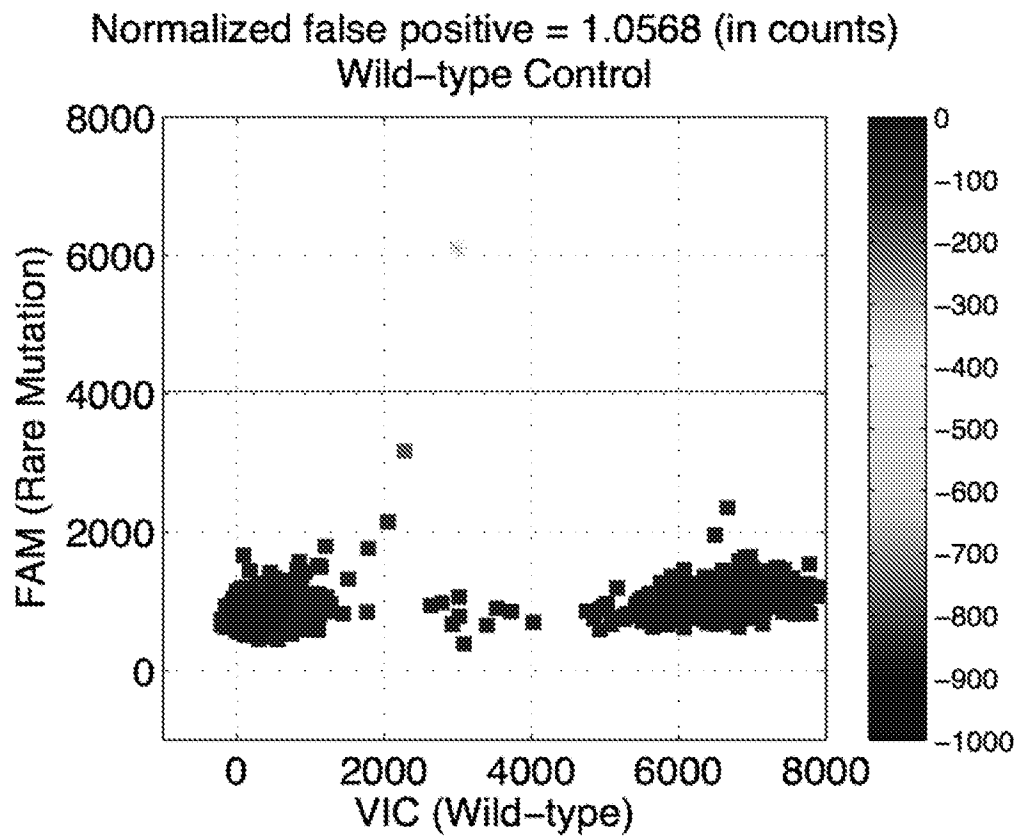


FIG. 2B

5/12

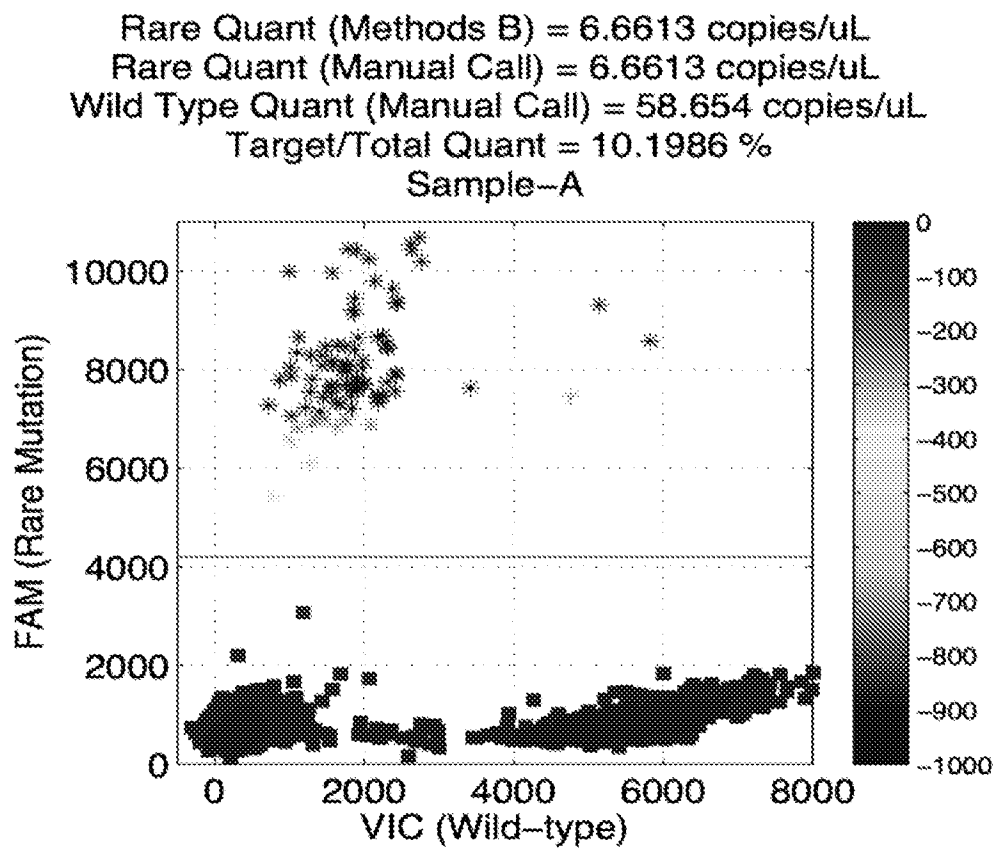


FIG. 3A

6/12

Rare Quant (Methods B) = 7.536 copies/uL  
Rare Quant (Manual Call) = 7.6735 copies/uL  
Wild Type Quant (Manual Call) = 59.0017 copies/uL  
Target/Total Quant = 11.3259 %  
Sample-A

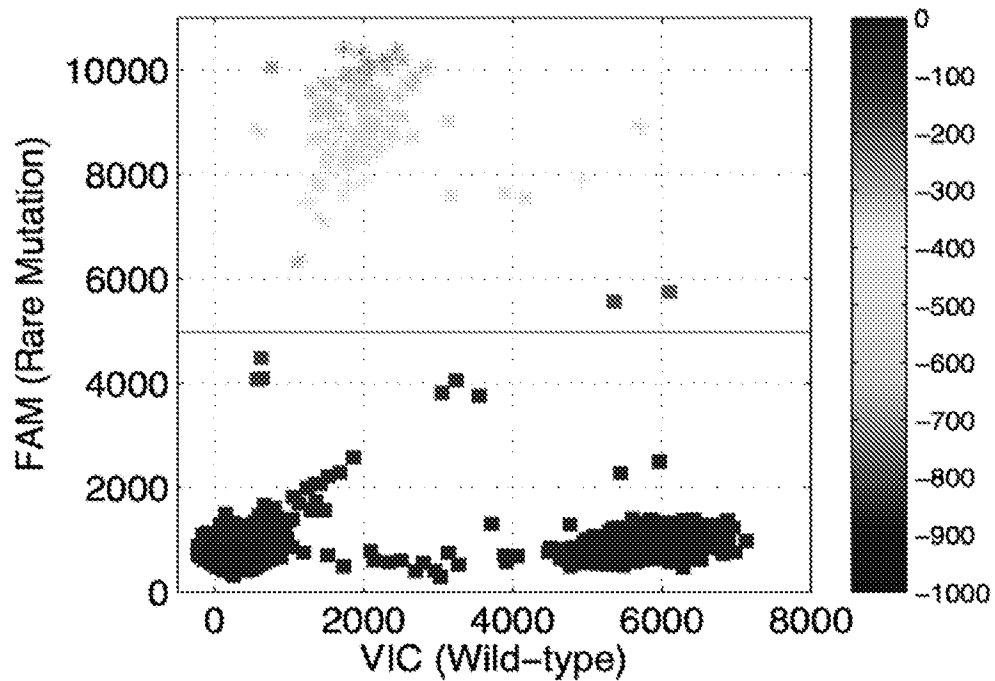


FIG. 3B

7/12

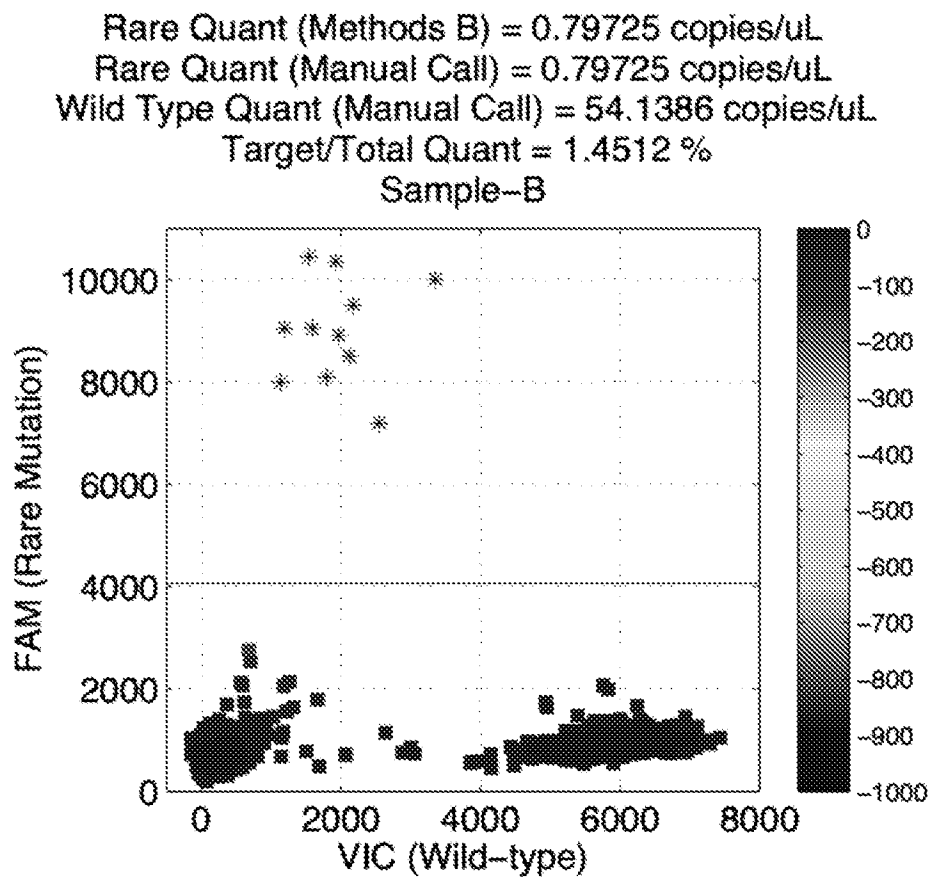


FIG. 4A

8/12

Rare Quant (Methods B) = 0.52991 copies/uL  
Rare Quant (Manual Call) = 0.52991 copies/uL  
Wild Type Quant (Manual Call) = 57.0315 copies/uL  
Target/Total Quant = 0.92061 %  
Sample-B

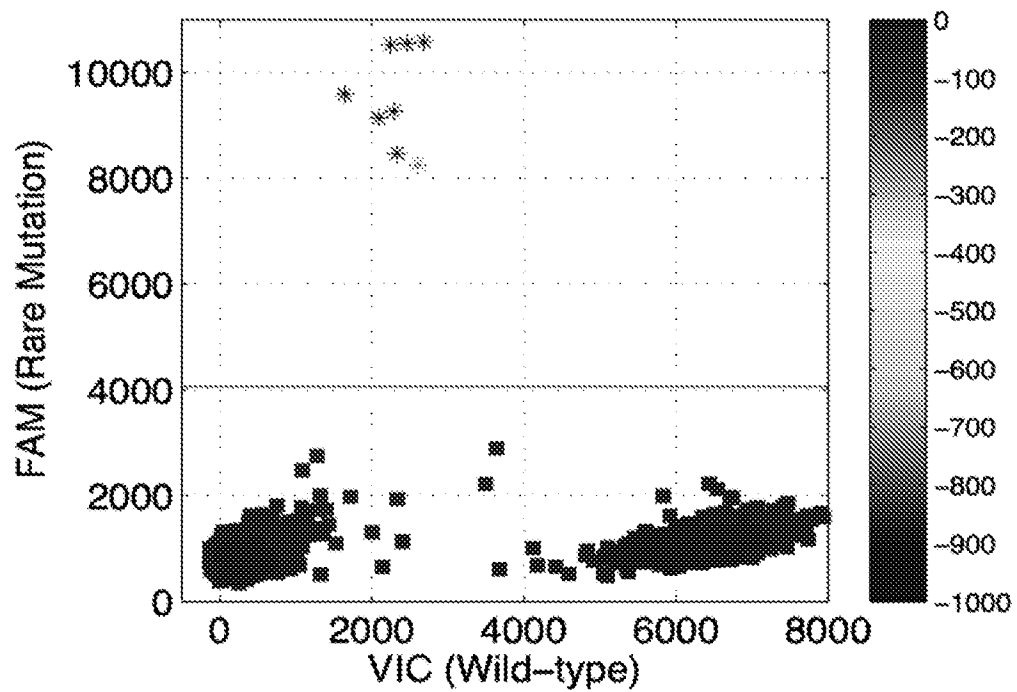


FIG. 4B

9/12

Rare Quant (Methods B) = 0.45414 copies/uL  
Rare Quant (Manual Call) = 0.45414 copies/uL  
Wild Type Quant (Manual Call) = 256.5344 copies/uL  
Target/Total Quant = 0.17672 %  
Sample-C

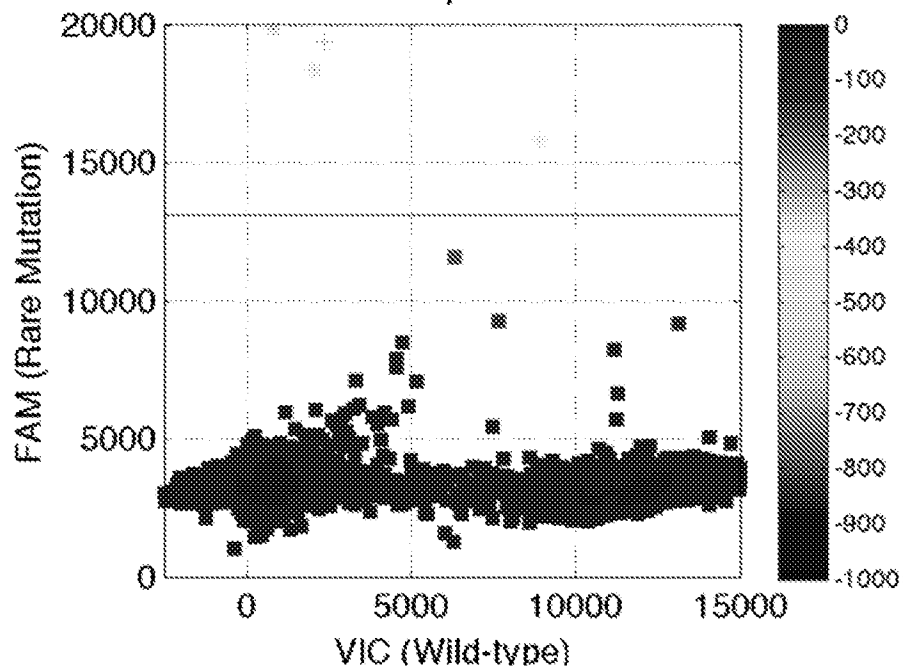


FIG. 5A

10/12

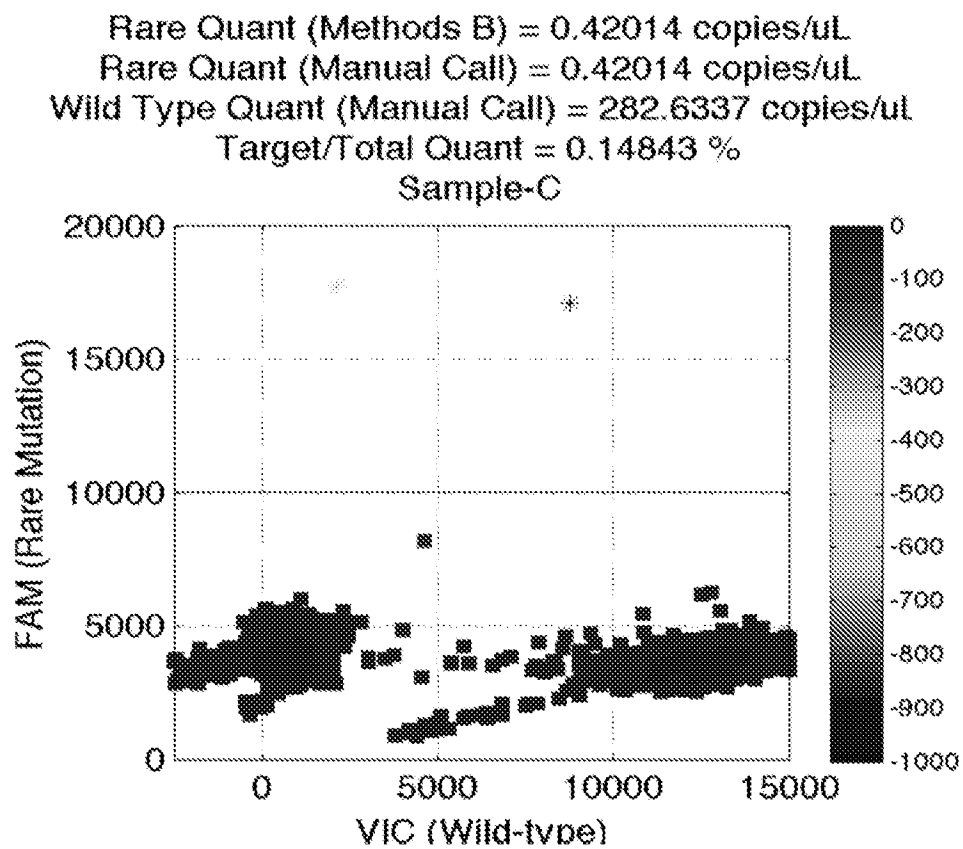


FIG. 5B

11/12

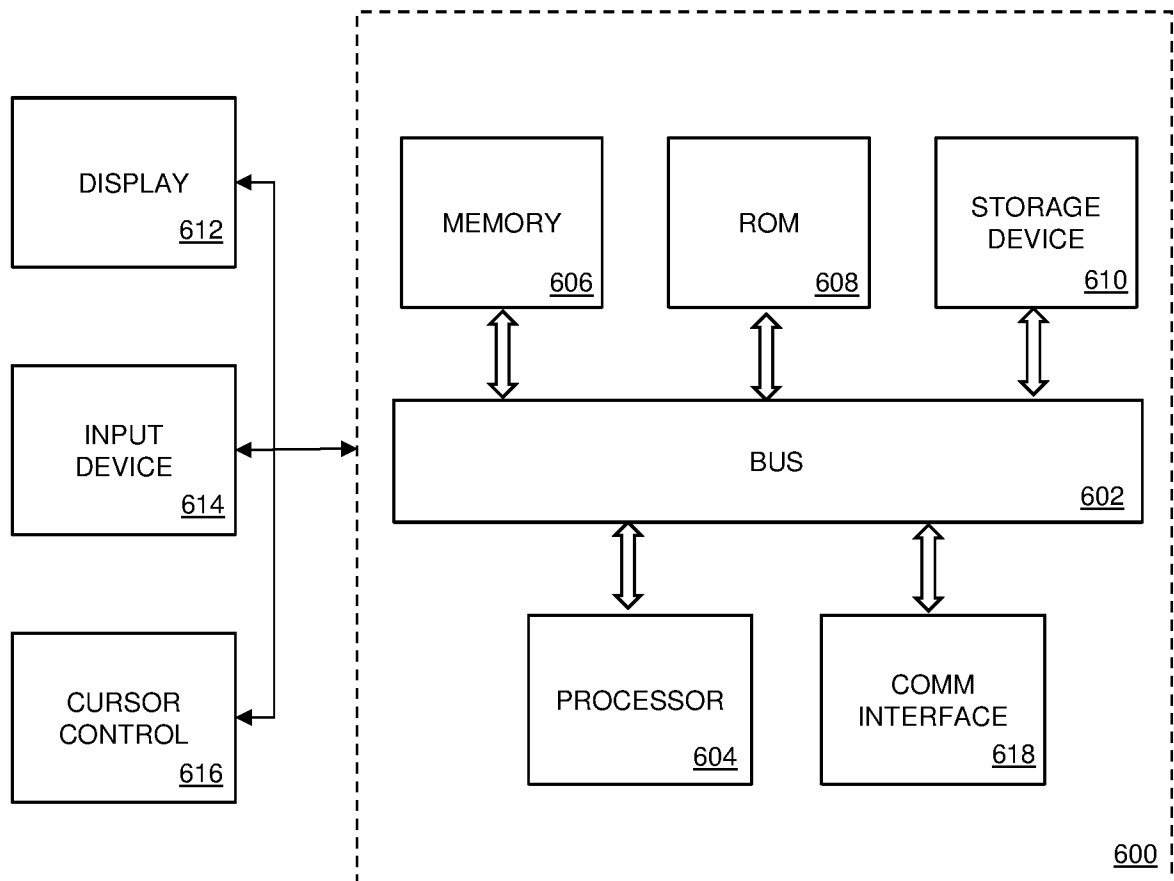


FIG. 6

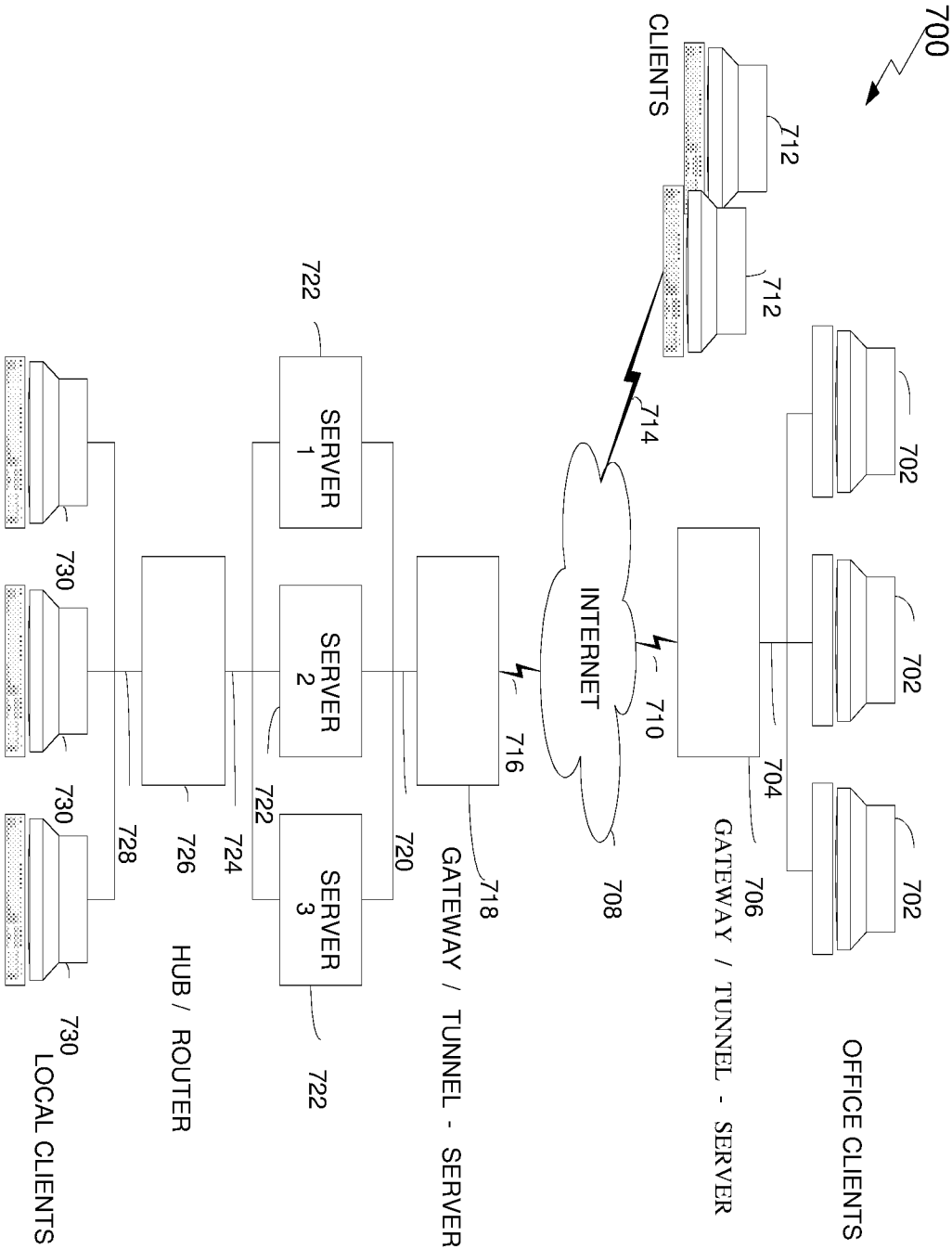


FIG. 7

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2016/035870A. CLASSIFICATION OF SUBJECT MATTER  
INV. G06F19/24  
ADD. G06F19/20

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                                                                                                                           | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | <p>Raúl Rojas: "The Secret Life of the Covariance Matrix",<br/>31 January 2009 (2009-01-31), XP055296076,<br/>Retrieved from the Internet:<br/>URL: <a href="http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/secretcovariance.pdf">http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/secretcovariance.pdf</a><br/>[retrieved on 2016-08-17]<br/>abstract<br/>page 4, paragraph 1 - paragraph 2<br/>page 5, paragraph 3 - paragraph 8<br/>page 8, paragraph 6<br/>-----<br/>-/-</p> | 1-12                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

## \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search

18 August 2016

Date of mailing of the international search report

29/08/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

Schmitt, Constanze

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2016/035870

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                                                               |                       |
|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                            | Relevant to claim No. |
| A                                                    | <p>COREN A. MILBURY ET AL: "Determining lower limits of detection of digital PCR assays for cancer-related gene mutations", BIOMOLECULAR DETECTION AND QUANTIFICATION, vol. 1, no. 1, 1 September 2014 (2014-09-01), pages 8-22, XP055296096, ISSN: 2214-7535, DOI: 10.1016/j.bdq.2014.08.001 abstract<br/>page 10, column 2, paragraph 1<br/>page 12, column 2, paragraph 7<br/>page 11, column 1, paragraph 2<br/>-----</p> | 1-12                  |