Cartagenia Bench Lab Case Study

Identifying Somatic Tumor-Only Variants with Intelligent Variant Filtration Strategies in Cartagenia Bench Lab NGS

At a Glance

In this case study, you will learn:

- How UHN has developed a custom variant filtration strategy in Cartagenia Bench Lab NGS to improve the identification of somatic tumor-specific variants in tumor-only testing, removing the need to sequence a reference sample.
- How building an internal knowledge base in Cartagenia Bench Lab NGS can aid in somatic- germline variant assessment.

Authors:

Tracy Stockley[1], PhD, FCCMG, FACMG

Mahadeo Sukhai [1], PhD

Suzanne Kamel-Reid [1], PhD, FACMG

Tina Smets[2], Msc

1. Genome Diagnostics, University Health Network, Toronto, Canada
2. Agilent Technologies, Leuven, Belgium

Introduction

Tumor molecular profiling by next-generation sequencing (NGS) aims to use information on somatic tumor-specific variants to direct cancer treatment. A significant challenge in tumor molecular profiling is identifying tumor-specific variants in the presence of germline variants which are also identified during tumor testing. The majority of germline variants are benign changes not contributing to the cancer occurrence, although rare germline cancer predisposition variants may also be identified during the testing of tumor tissue.

In order to identify true somatic tumor-specific variants, a common approach is the sequencing of both tumor and normal tissue (often peripheral blood), with variants from each tissue source used to classify variants as either somatic or germline. However paired tumor-normal testing leads to increasing costs due to the need to test both tissue sources. In addition paired tumor-normal testing may identify potentially unwanted incidental germline cancer predisposition variants.

Ideally, a solution is needed that can improve the identification of somatic tumor-specific variants obtained from NGS testing of tumor tissue only, without the need to sequence normal tissue for comparison to germline variants. We sought to evaluate the use of Cartagenia Bench Lab NGS for this purpose, using 1120 solid tumor cases for which paired tumor-normal NGS variant results were available at the Genome Diagnostics Lab at the University Health Network (UHN; Toronto). Using a custom tree designed using the Cartagenia Bench Lab NGS, these cases were assessed for correct classification of somatic tumor-only variants.

UHN is a research and healthcare network of four major hospitals in Toronto, including Toronto General Hospital and Princess Margaret Cancer Centre, and is affiliated with the University of Toronto. The scope of research and complexity of cases at UHN have made it a national and international source for discovery, education and patient care.

Approach

In this study variants from 1120 solid tumor cases profiled using a NGS panel assessing regions of 48 cancer-related genes (Illumina TruSeq Amplicon – Cancer Panel) were examined. Variants were processed by use of MiSeq Reproter (Illumina). NGS profiling was performed on DNA from both tumor and normal tissue (peripheral blood leukocytes) for each case.

A custom tumor-only somatic variant classification tree was designed in Cartagenia Bench Lab NGS (see Figure 1). The tree included annotation filters and settings designed to yield the maximum number of true somatic variants, while simultaneously reducing the number of germline variants inadvertently classified as somatic variants. The custom Cartagenia Bench NGS variant tumor-only classification tree was initially used in a pilot study of 100 cases, with refinement of tree based on pilot data prior to use in the total 1120 cases.

Tumor-only Somatic Variant Classification Tree

In the first stage, all variants derived from tumor samples are filtered using quality criteria. The quality criteria assessment includes assessment of quality of NGS data and read depth, with only variants making it through the quality PASS filter (i.e. those with high-confidence base calls and no strand bias) and with a read depth of at least 250x included in the next steps. A third quality criterion was applied using a Variant Allele Frequency (VAF) threshold, with variants less than the 5% VAF threshold excluded due to known frequent occurrence of low-level (VAF<5%) variant artifacts in the methods used in this study.

In the second stage, the variants in the predicted somatic category were filtered to focus on clinically relevant variants. Variants were excluded if they were present in introns (intronic variants within 2 bp of exons retained for consensus splice site assessment), were synonymous or were known recurrent artifacts based on previous analyses.

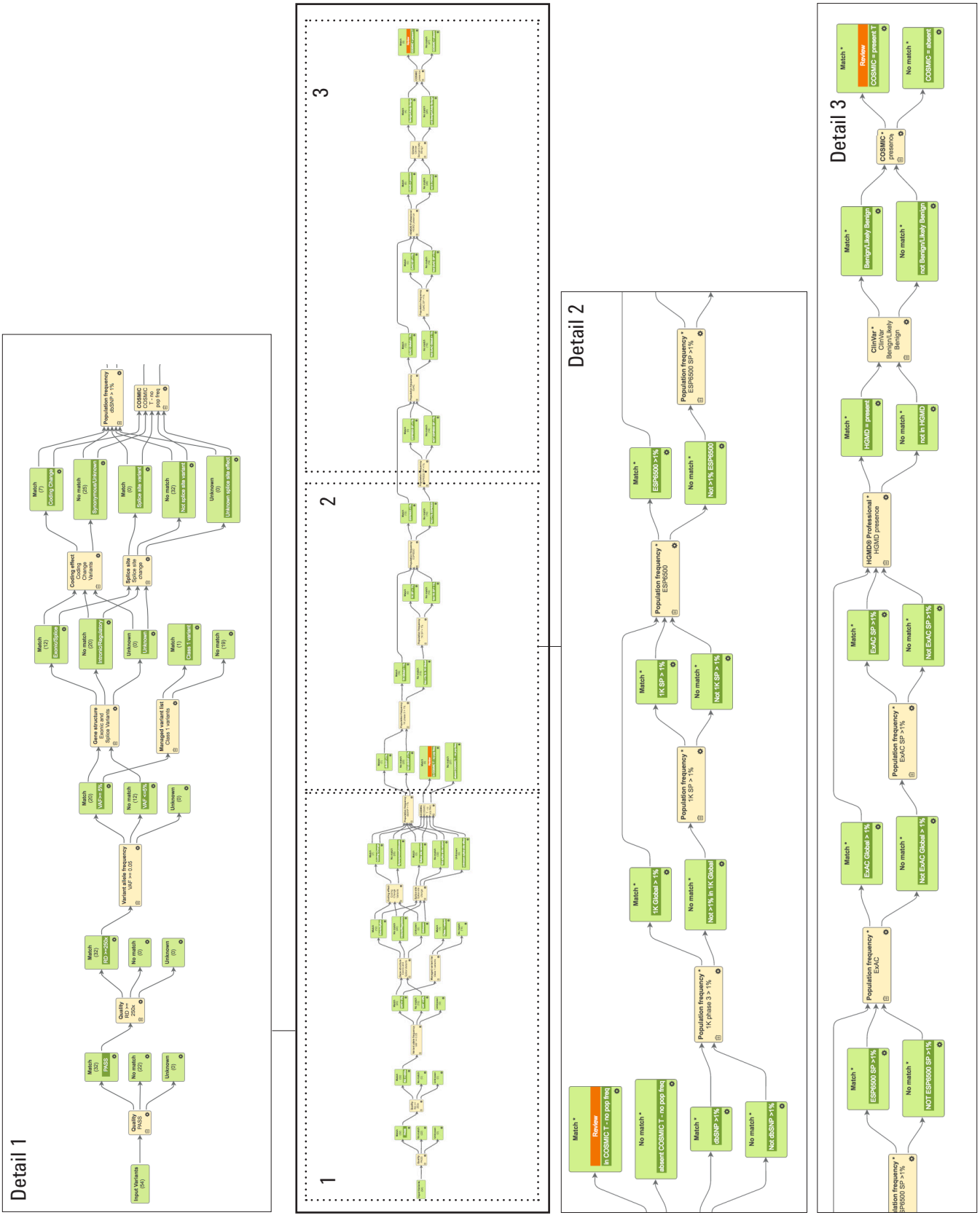


Figure 1: Overview of custom tumor-only somatic variant classification tree designed to yield the maximum number of true somatic variants by means of a flexible labeling strategy. Detail 1-2-3: All variants derived from tumor samples are filtered to meet the desired quality criteria (PASS, read depth of at least 250x) and of these variants only the ones with a VAF score above 5% are considered to be potentially somatic. To decipher the clinical relevance of variants filters relating to coding effect, population frequency and the in-house knowledge base or Managed Variant List (MVL) are applied while presence in clinically relevant databases is evaluated. Efficient labeling enables high-throughput categorization of variants (green) and in-depth review (red).

In the third stage, the variants passing the described quality filters were subsequently checked against a number of population frequency databases, with Minor Allele Frequency (MAF) at a threshold of 1%. Variants present in any of these databases, either at the population or subpopulation level, were transitioned to a predicted germline category. For these predicted germline variants, the presence of any known somatic variant was evaluated using a lab-specific knowledge base or Managed Variant List (MVL) of known recurrent somatic variants with clinical importance. Using the MVL assessment, any somatic variants present in the MVL but inadvertently sorted into the predicted germline list would be transferred to the predicted somatic category. To supplement this automated approach towards predicting the somatic or germline entity of a variant, a broad range of other annotation filters is applied. For example, presence in clinically relevant databases Human Gene Mutation Database (HGMD), ClinVar or COSMIC can be applied to the predicted somatic category to further differentiate germline variants from somatic tumor variants. Thanks to Bench Lab's flexible labeling functionality any additional annotation information can easily be included in variant assessment and consulted in the variant review tab (Figure 2).

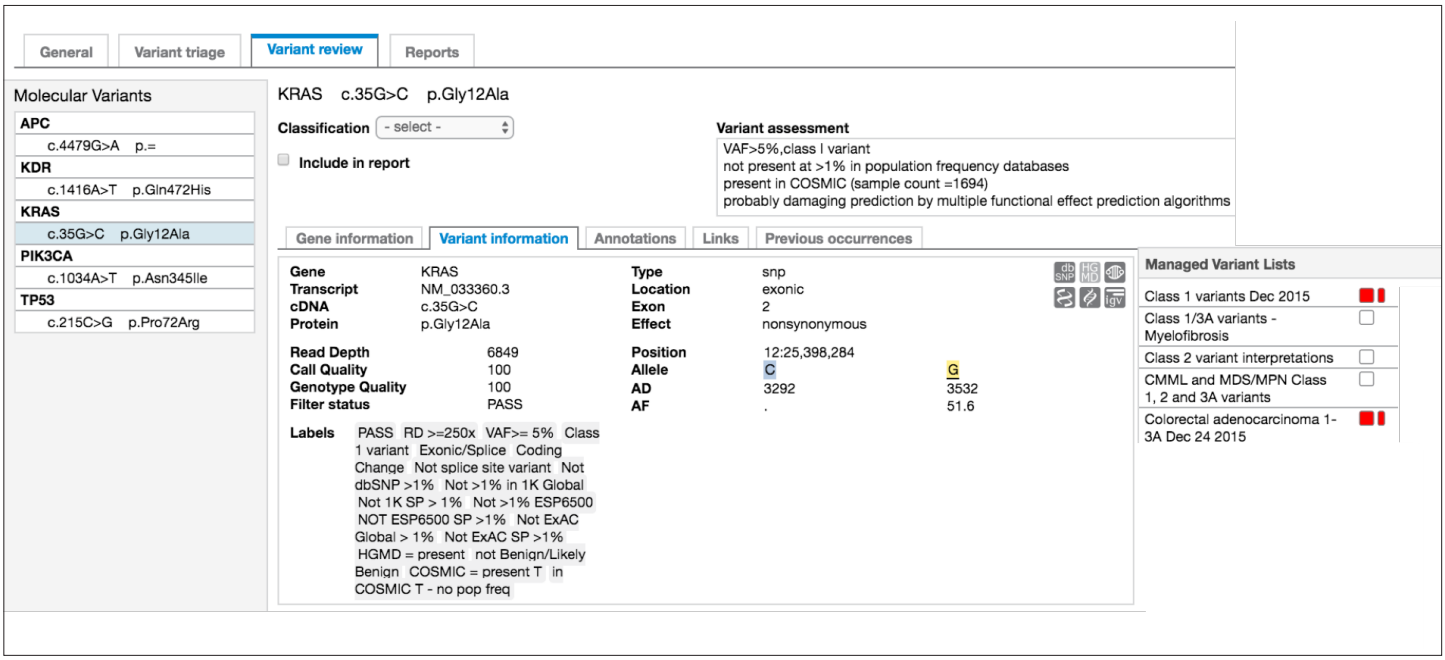


Figure 2: In the variant review tab all information regarding a certain variant is brought together, providing direct access to the variant's annotations, attached labels and the lab's knowledge base or MVLs. For the KRAS p.Gly12Ala variant we observed a VAF above 5%, presence in COSMIC (n =1694 samples), a negative prediction by multiple functional effect prediction algorithms and absence at a frequency below 1% for all population frequency databases.

In addition to predicting whether a variant would fall into the germline or somatic category, the custom tree could also be designed to identify known germline pathogenic variants that are potential inherited germline cancer predisposition variants. The inclusion criteria for pathogenic germline predisposition cancer variants MVL could be the following: VAF between 35% and 65%, association with gene known to be correlated with cancer predisposition (for this panel, relevant genes were APC, PTEN, RB1, RET, MLH1, STK11, TP53, VHL) and present as pathogenic or likely pathogenic variant in ClinVar (Figure 3). This addition to the tree would be important for settings in which both identification of somatic variants and germline cancer predisposition variants is desired.

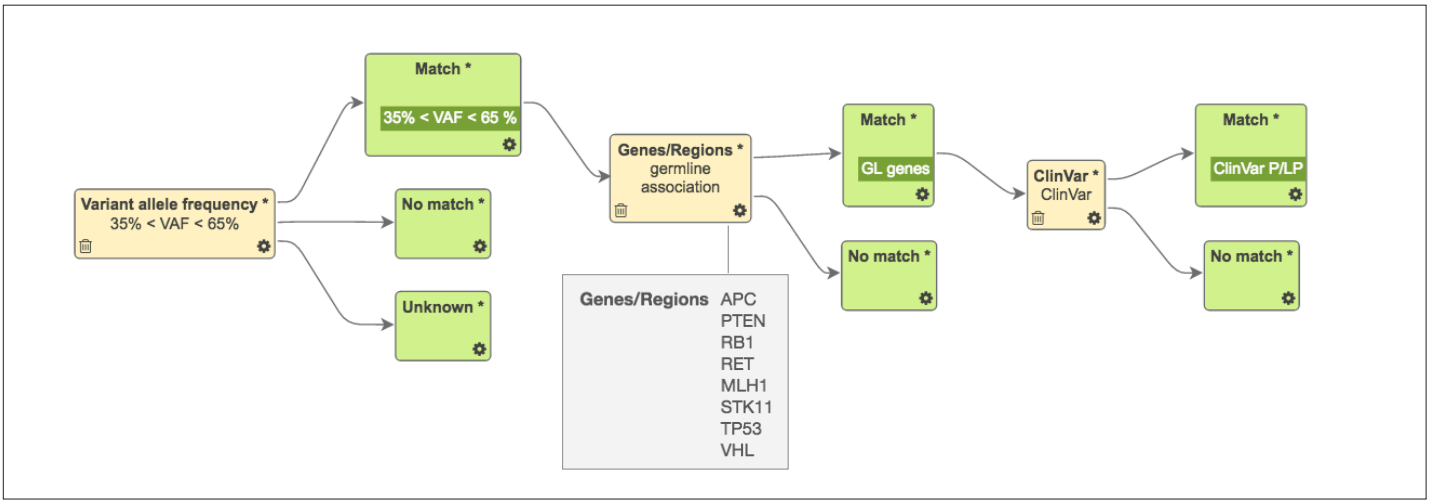


Figure 3: Identification of potentially inherited germline cancer predisposition variants by selecting the variants with a VAF between 35% and 65%, situated within the gene list and present as pathogenic or likely pathogenic in ClinVar. This addition to the tree made it possible to simultaneously build an MVL for germline cancer predisposition variants.

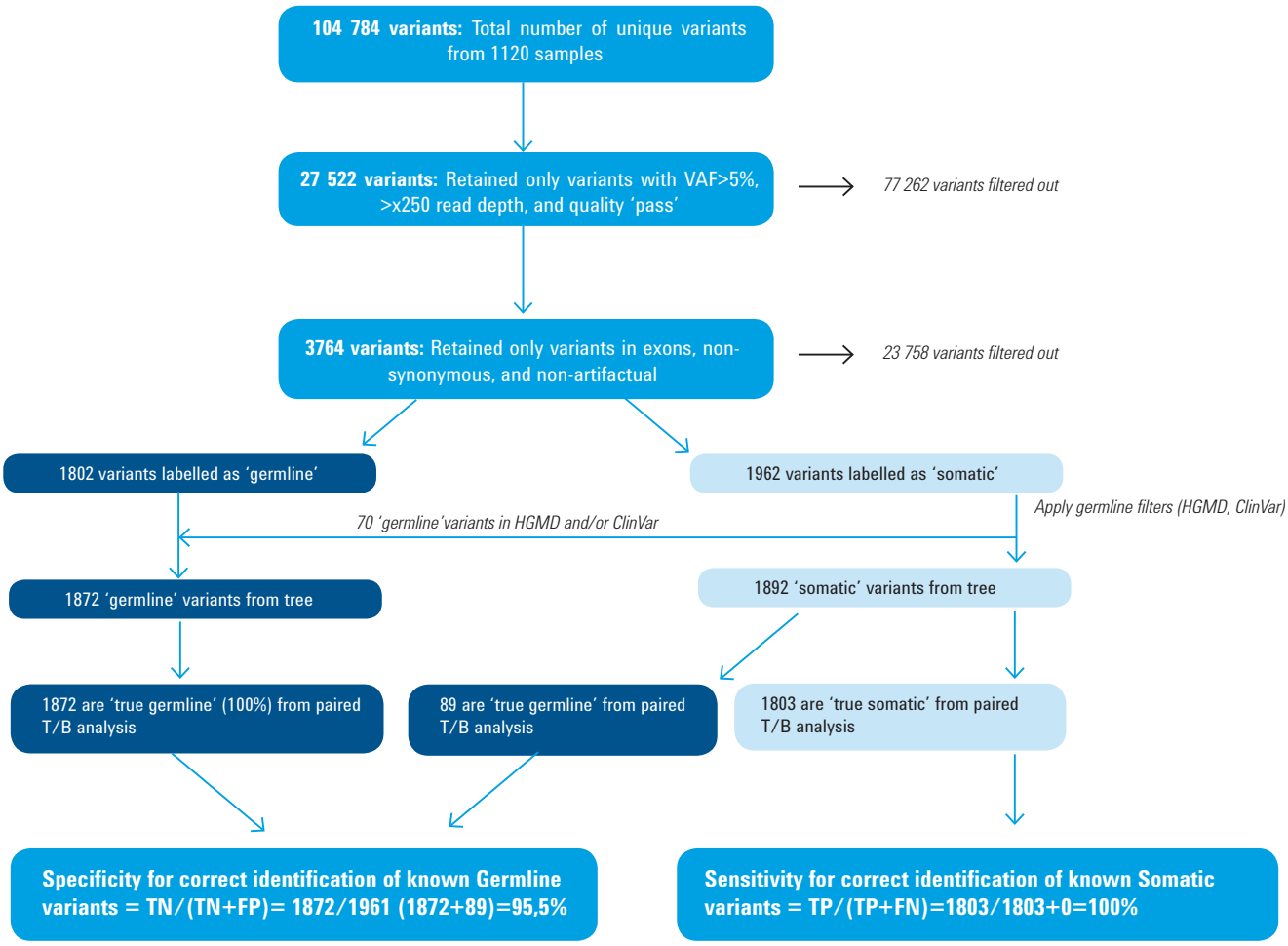
Results and Comparison of Tumor-Only Variant Classification to Paired Tumor-Normal Classification

Results of application of the custom Cartagenia Bench NGS tumor-only somatic variant classification tree to 1120 cases is shown in Figure 4. Overall 104,784 unique variants were identified in NGS analysis of 1120 tumor tissue samples, of which 27,522 variants were retained after quality filters, and 3764 variants retained after clinically relevant filters were applied. Of these, 1802 were labelled as 'germline' and 1962 as 'somatic' variants by the tumor-only tree shown in Figure 1. For the 1962 variants labelled somatic, a further filter of presence in HGMD and/or ClinVar was applied which allowed reclassification of 70 variants as germline, for final counts of 1892 variants labelled as somatic and 1872 variants labelled as germline by the tree. Of the 1872 variants labelled germline, all 1872 were also identified as true germline by comparison to the paired tumor-normal analysis, while for the 1892 variants labelled as somatic 1803 were true somatic and 89 were true germline from the paired tumor-normal analysis. Based on these values, the final sensitivity of the tumor- only somatic variant tree to correctly identify known somatic variants was therefore 100% (1803/1803) and specificity for exclusion of known germline variants from an incorrect labeling as a somatic variant was 95.5% (1872/1961).

Conclusion

The use of a custom variant filtration strategy in Cartagenia Bench Lab NGS to identify tumor-specific variants from tumor-only testing can detect true somatic variants with a sensitivity of 100% and a specificity of 95% as compared to use of paired tumor-normal analysis. As such, the usage of an optimized variant filtration pipeline in Cartagenia Bench Lab NGS removes the need for running a normal reference sample which greatly reduces the cost, effort and hands-on time. Additionally, building an internal knowledge base in Cartagenia Bench NGS can aid in the assessment of molecular variants in somatic and inherited cancer applications.

Identifying Somatic Tumor-Only Variants with Intelligent Variant Filtration Strategies



LEARN MORE:
www.agilent.com/lifesciences/cartagenia

Email us at
cad.info@agilent.com

Trusted Answers. Together.

PR7000-0283

Printed in USA
Original publication date November 01, 2016
Revision Date April 28, 2017
5991-8026EN

Cartagenia Bench Lab is marketed in the USA as exempt Class I Medical Device and in Europe and Canada as a Class I Medical Device.

Agilent Technologies