**Next-Generation Sequencing Infrastructure Policies**

1. **PRINCIPLE:**
   1. The following policies and procedures form the infrastructure for quality management of all Next-Generation Sequencing (NGS) Assays and Procedures.
2. **Exception Log:**
   1. The laboratory maintains an exception log (Excel file) for patient specimens and/or NGS assay runs for which steps deviate from written procedures.
      1. The log is located in the laboratory’s shared network drive under \\lsfile03\RICMBLAB$\Problem log\NGS
   2. The exception log should indicate the specific Soft Molecular worksheet or tasklist number if the exception applies to a whole run or individual patient sample numbers. In the case of worksheets and tasklists, these can be used to determine the patient samples that were included in the run through the assay worksheet or lab information system.
      1. The log will include each deviation and the reason(s) for the deviation.
      2. The Laboratory Director or designee(s) will review these records and comment on any issues or corrective actions taken as a result of these reviews.
3. **Data Management:**
   1. The laboratory must ensure that patient confidentiality, security, and data integrity are maintained during internal and external storage and transfer of data (including NGS data).
      1. This must be in compliance with Lifespan IS policies, which can be obtained from the Lifespan Intranet, under Policies>Lifespan Corporate Services > Information Services
   2. Maintaining confidentiality during transfer of data
      1. For detailed Lifespan IS policies, refer to the Transmission Security Policy (HSP-97).
      2. Copies of HIPAA Business Associate Agreements for referral laboratories or companies storing datasets will be maintained by the Purchasing Department.
      3. A Data Transfer Log will be maintained on the network drive under \\lsfile03\RICMBLAB$\Problem Log\DataManagement
         1. Information logged may include, as applicable: data files sent, date/time, sender, method of sending, and receiver.
      4. Transfer of data should be in compliance with Lifespan IS Policy HSP-97. Specific examples of transfer methods include:
         1. Illumina:
            1. Any log files that do not contain PHI may be sent as necessary, according to Lifespan security policies.
            2. If sample data must be sent, save all files in an encrypted \*.zip file (in accordance with Lifespan IS policies).
            3. Send the file via a secured share file.
            4. If this is not possible, an encrypted email may be sent (using “PHI” in the email subject), according to Lifespan security policies.
         2. Pillar:
            1. Pillar Variant Analysis Toolkit (PiVAT) error logs do not contain PHI and may be sent as necessary, according to Lifespan security policies.
         3. ArcherDX:
            1. If sample data must be sent, save all files in an encrypted \*.zip file if possible (in accordance with Lifespan IS policies).
            2. Send the file via a secured share file.
            3. If this is not possible, an encrypted email may be sent (using “PHI” in the email subject), according to Lifespan security policies.
      5. To ensure confidentiality of patient data, the Molecular Genomic Pathology Lab will follow Lifespan IS policies regarding data encryption, use of secure and encrypted protocols for electronic data transfer (e.g. SFTP, HTTPS, FTPS), system and user authentication, activity logs, access restrictions, and appropriate data backups.
      6. Records of NGS data transmission and storage are maintained as described in the appropriate sections of this document.
   3. Storage of files, type of data to be retained, and length of retention:
      1. Data necessary to re-review cases as originally performed for original results will be retained indefinitely on the RICMBLAB$ shared drive and/or the lab information system.
         1. This includes specimen tracking and quality metrics data, sequencing run quality metric reports, log or configuration file information regarding bioinformatics pipeline parameters, sequence read alignments, exception log information, variants undergoing manual review, and files containing filtered and/or interpreted variants.
         2. The retained files and records must also be structured to facilitate inter-laboratory replication of the original analyses, annotations and/or interpretation, whether initiated by the laboratory or at the request of the referring physician or patient tested.
      2. Primary NGS data:
         1. All raw files are initially maintained on the instrument.
         2. Raw run files are then transferred to the [\\LSFILE14\MGPGenomicData$](file:///\\LSFILE14\MGPGenomicData$) shared drive for a minimum of 6 months to support primary results generated and re-analysis.
         3. A subset of primary data, including FASTQ files and sequencing quality controls information (such as the data generated by Illumina SAV software), is copied to the corresponding directory for each run under Lifespan RICMBLAB$ network drive, which is backed-up routinely as per Lifespan IS policies, and stored for a minimum of 2 years. If runs need to be re-analyzed, these files (along with the pipeline version information stored in the assay run worksheets) can be used to perform the re-analysis.
         4. Additional MiSeqDx and NextSeq run data:
            1. Run-level data should be transferred from the instrument to the Lifespan network drive according to each assay Procedure.
      3. Secondary NGS data (the data generated by bioinformatics pipelines):
         1. Actionable Mutation Panel 2
            1. PiVAT data:

Data is generated and exported from the software according to the AMP2 assay Procedure.

All analysis data is stored on the RICMBLAB$ network drive, which is backed-up per Lifespan IS policies. This includes BAM, VCF, and result files, as generated by the vendor software.

Prior data is removed from the virtual machine which runs the vendor software.

* + - * 1. Lifespan AMP Reporter 2 data:

Data is generated on the Lifespan Linux-based server, then transferred to the RICMBLAB$ network drive. All prior run data is deleted from the server.

All analysis data stored on the RICMBLAB$ network drive is backed-up per Lifespan IS policies. This includes BAM, VCF, intermediate files, summary files, and other metafiles, as generated by the custom-made pipeline.

Security measures in Linux server: All programs in LAR pipeline are installed and run on the Linux server (lsmplinux3) located in the Lifespan IS department. None of the programs in the pipeline are cloud-based. This server is on the Lifespan network and all security patches have been installed and updated by the Lifespan IT department. Logins on the server are strictly password-protected. Authorized personnel have separate usernames and passwords to log into the server. The genomic data analyzed on the Linux server is coded and only laboratory-defined sample IDs are used in the analysis.

* + - 1. Solid Tumor Fusion Panel
         1. Archer Analysis data:

Data is generated using a Virtual Machine using VMWare ESXi. The host server is located within and maintained by the Lifespan IS department.

A subset of the data is downloaded from the software according to the Solid Tumor Fusion Assay Procedure.

All downloaded files are stored on the RICMBLAB$ network drive, which is backed-up per Lifespan IS policies. This includes BAM, VCF, and summary files, as generated by the vendor software.

Prior run data may be maintained on the virtual machine hosted within the Lifespan server.

When the database is full, since all required run data has already been exported, data can be deleted with Director of Clinical Bioinformatics’ approval.

* + - * 1. Lifespan Fusion Reporter data:

Data is generated on the Lifespan Linux-based server, then transferred to the RICMBLAB$ network drive. All prior run data is purged from the server.

All analysis data stored on the RICMBLAB$ network drive is backed-up per Lifespan IS policies. This includes BAM, VCF, intermediate files, summary files, and other metafiles, as generated by the custom pipeline.

Security measures in Linux server: All programs in the bioinformatics pipelines are installed and run on the Linux server (lsmplinux2) located in the molecular laboratory. None of the programs in the pipeline are cloud-based. This server is on the Lifespan network and all security patches have been installed and updated by the Lifespan IT department. All software is installed and maintained by Lifespan IS. Logins on the server are strictly password-protected. Authorized personnel have separate usernames and passwords to log into the server. The genomic data analyzed on the Linux server is coded and only laboratory-defined sample IDs are used in the analysis.

* + 1. Tertiary Data:
       1. Actionable Mutation Panel 2
          1. Lifespan Variant Review Tooldata:

All files generated by this software (MS Excel files), are stored in the appropriate RICMBLAB$ network drive test run folder, according to the Actionable Mutation Panel 2 Procedure. These files/folders are backed-up per Lifespan IS policies.

* + - * 1. Lifespan Variant Review ToolSummary File

An Excel file is generated with each run summarizing the variants identified by the pipeline, as well as interpretation by the Bioinformatician or designee. This file will be maintained in the RICMBLAB$ network drive and backed-up, as described above.

* + - 1. Solid Tumor Fusion Panel
         1. Fusion Review Tool data:

All files generated by this software (MS Excel files), are stored in the appropriate RICMBLAB$ network drive test run folder, according to the Fusion Panel Procedure. These files/folders are backed-up per Lifespan IS policies.

* 1. Transfer of Data from the NGS Instruments for long-term storage
     1. Data from the instruments can be transferred to the [\\LSFILE14\MGPGenomicData$](file:///\\LSFILE14\MGPGenomicData$) shared drive or a Lifespan-supplied external hard drive with real-time EAS encryption and a padlock.
     2. A log of data Long Term Storage events will be kept on the network drive under [\\lsfile03\RICMBLAB$\DataManagement](file:///\\lsfile03\RICMBLAB$\DataManagement).
        1. Initially, files were transferred to an external hard drive and then subsequently moved to the Lifespan network drive. These activities were logged.
        2. Subsequently, files were directly transferred from the instrument to the network folder, which will not be logged as it is part of routine processes.
        3. However, when files are deleted from long-term storage, the events will continue to be logged.
  2. Transfer of Data from a non-networked computer to a networked computer:
     1. Data can be transferred from the non-networked computer to a Lifespan-supplied encrypted USB flash drive and external hard drive with real-time EAS encryption and a padlock.
        1. Data is then transferred from the flash drive or external hard drive to the Lifespan network drive for long-term storage.
     2. Alternatively, data can be transferred through mapped network drives.
        1. On the MiSeqDx, move files to the SharedData folder.
           1. On the Lifespan computer that is connected to the MiSeqDx, map the network drive [\\M-M70291R\SharedData](file:///\\M-M70291R\SharedData)
           2. Then, move all files (as appropriate) from the SharedData folder to the [\\LSFILE14\MGPGenomicData$](file:///\\LSFILE14\MGPGenomicData$) shared drive.
        2. On the NextSeq, files should be located in the mapped network drive nextseq\Runs (hosted on lsmplinux2).
           1. On a Lifespan computer, map the nextseq\Runs folder.
           2. Then, move all files (as appropriate) from the Runs folder to the [\\LSFILE14\MGPGenomicData$](file:///\\LSFILE14\MGPGenomicData$) shared drive.

1. **QUALITY MANAGEMENT PROGRAM:**
   1. The laboratory has a Quality Management (QM) program for the NGS analytical wet bench process.
      1. The laboratory records all deviations from the standard operating procedure and assesses their impact on quality of the clinical assay.
         1. A paper log of instrument and assay-related problems, with their resolutions, is kept near the instrument.
         2. In addition, an electronic log of all procedure deviations is kept on the network drive under G:\Problem log\NGS\NGS\_ExceptionLog.xlsx.
         3. Corrective actions and exceptions are recorded for every step that fails performance standards, as set by the laboratory.
         4. All exceptions to standard operating procedures are approved by the laboratory director or designee.
      2. Specific controls, metrics, and quality control parameters are used to monitor and assess each analytical run.
         1. For the Actionable Mutation Panel 2:
            1. A log is kept of all runs and the Index IDs used for each sample, called “Index Codes Log”.
            2. In the same file, the following sequencing metrics are recorded with each run (as obtained from the Illumina Sequencing Analysis Viewer software):

% > Q30

Aligned % (R1)

Cluster Density

Cluster PF (%)

Reads (M)

Reads PF (M)

* + - * 1. Once per month, this log file will be reviewed by the Director to monitor for changes and trends in metrics.
      1. Individual Actionable Mutation Panel 2 runs:
         1. Runs should ideally meet the following metrics. Those that do not meet the following metrics should be interpreted appropriately by the Director:

Q30 > 80% (percent of bases with at least Q30)

% Clusters PF (Passing Filter) > 80%

* + - * 1. Additional metrics to review:

% aligned (ideally 4-10%, but may vary with run)

Cluster density (ideally 950-1,200K/mm2 per Illumina, but AMP runs tend to have lower values with successful data)

These metrics should be monitored but do not directly determine run pass/fail decisions)

* + - 1. Solid Tumor Fusion Panel
         1. A log is kept of all runs and the Index IDs used for each sample, called “Index Codes Log”.
         2. A separate log is maintained for NextSeq SAV run metrics for FusionST on the RICMBLAB$ shared drive. The following sequencing metrics are recorded with each run (as obtained from the Illumina Sequencing Analysis Viewer software):

% > Q30

Aligned % (R1)

Cluster Density (k/mm2)

Cluster PF (%)

* + - * 1. Once per month, this log file will be reviewed by the Director to monitor for changes and trends in metrics.
      1. Individual Fusion Panel runs:
         1. Runs should ideally meet the following metrics. Those that do not meet the following metrics should be interpreted appropriately by the Director:

Q30 > 75% (percent of bases with at least Q30)

% Clusters PF (Passing Filter) > 75%

* + - * 1. Additional metrics to review:

% aligned (ideally 20%; should be below 50%, but may vary with run)

Cluster density (ideally 170-220 K/mm2)

These metrics should be monitored, but do not directly determine run pass/fail decisions.

* + 1. Individual sample Quality Control
       1. The Actionable Mutation Panel 2 has two primary QC steps.
          1. The first step is an RT-PCR assay with two parameters:

For each sample, the Functional Copy score is determined and compared to a validated threshold for acceptability.

Samples which pass QC threshold show Functional Copies > 200.

An Inhibitor Assay helps determine if a PCR inhibitor may be present in the sample.

In the Box 2 Worksheet, review the “Inhibitor Assay Avg”.

This number is the average Ct of the Target 2 assay for each sample.

Any sample with a Target 2 Ct > 2 cycles above this number may exhibit signs of PCR inhibition. Review the sample accordingly and repeat testing under appropriate conditions, if necessary.

* + - * 1. The second QC step quantifies the library and determines if the library can be added to the pool for NGS.

See the AMP2 procedure for details on these criteria and the steps to take if a sample does not meet the appropriate thresholds.

* + - 1. Solid Tumor Fusion Panel
         1. The PreSeq assay is used for initial assessment of sample quality but is not used to determine sample pass/failure.
         2. Archer Analysis QC should be used to determine sample acceptability, especially when no Strong fusion calls are seen.
         3. However, if a Strong fusion call is made in a sample that fails QC, the fusion may be reported at the discretion of the pathologist with an interpretive comment about the QC results. Results should be interpreted with caution. If appropriate, testing may be repeated for the sample.
  1. The laboratory has a Quality Management (QM) program for the Bioinformatics workflow/pipeline.
     1. The laboratory records all deviations from the standard operating procedure and assesses their impact on quality of the clinical assay, as described above (NGS Exception Log).
     2. Metrics and QC parameters for Actionable Mutation Panel 2 include:
        1. Percent of bases covered at specific read depths are determined by PiVAT.
        2. The percentage of reads aligned in pairs is determined by PiVAT and reviewed by the Director of Clinical Bioinformatics.
        3. For somatic cancer assays, monitoring of limit of detection controls for determination of assay sensitivity over time (e.g., weekly, monthly, quarterly)
           1. For the Actionable Mutation Panel 2: with each run, a HD701 cell line derived control is tested with a variety of substitutions and indels in different genes; some variants are near the LOD of the assay. This control allows determination of test reproducibility (e.g., identification of the same variants in a specimen) over time.
           2. For the Solid Tumor Fusion Panel: with each run, a sensitivity control derived from a cell line, or a previous positive sample is used and this sample has a fusion call with metrics near the thresholds for positivity.
        4. Sample Quality Control checks (PiVAT)
           1. For each sample, average coverage of targeted bases is determined and compared to a reference average that is determined during assay validation.

For each amplicon, the coverage must be > 200x; amplicons with lower coverage fail the QC threshold.

* + - * 1. Criteria for Failed state:

greater than 20% of the targeted bases have less than 100x coverage

* + - 1. Sample QC checks (AMP Reporter)
         1. FastQC is a program designed to perform quality control checks on raw sequencing data. It provides average quality scores for every base in each read. In addition, there are graphs indicating run metrics, including a tile graph showing quality of the reads in the results. FASTQ files produced by the MiSeqDx are used as input.

Samples pass QC if the majority of the reads have average quality scores >20.

* + - * 1. Any sample that does not pass QC should be reviewed and reported as insufficient, repeated, or interpreted with caution at the Director’s discretion.
    1. Run QC is determined using Illumina’s Sequencing Analysis Viewer software, as described above.
    2. See the Actionable Mutation Panel 2 and Solid Tumor Fusion Panel Procedures for more details about Quality Control procedures and steps to take when samples/run do not meet established criteria.
  1. The specific version(s) of the assay and bioinformatics process (pipeline) used to generate NGS data files are traceable for each patient report.
     1. For the Actionable Mutation Panel 2, each run directory has an Excel file (AMP2\_Software\_versions) which contains a list of all programs and reference files with associated version numbers used during the assay run.
        1. This can be used to perform tracebacks and recreate data, if necessary.
        2. Recorded version numbers include:
           1. PiVAT
           2. Lifespan AMP Reporter 2

Given a single version number as a whole.

Version numbers of individual components

Reference genomes

* + 1. Lifespan Variant Review Toolfor the Solid Tumor Fusion Panel, the Soft Tasklist contains a list of all software packages, scripts, and databases with associated version numbers used during the assay run.
       1. This can be used to perform tracebacks and recreate data, if necessary.
       2. Recorded version numbers include:
          1. Archer Analysis
          2. Lifespan Fusion Reporter
          3. Fusion Review tool
    2. Any changes to software packages, scripts, databases, configuration files or other pipeline components require tracking in the pipeline version control system and updating to a new version.
  1. Lab Records
     1. The methods, instruments, and reagents used throughout the entire testing process for a given sample (or batch of samples) are recorded in the assay-specific Excel worksheets for each sample or in the lab information system.

1. **MONITORING UPGRADES:**
   1. Procedure for monitoring, implementing, and recording upgrades to instruments, sequencing chemistries, and reagents or kits used to generate NGS data.
      1. Assays have been validated using specific versions of reagents, sequencing chemistries, and associated software. Therefore, no upgrades shall occur without appropriate review and, if necessary, validation. However, the lab shall periodically monitor for the availability of updates/upgrades.
         1. Monitoring activities for the AMP 2 and Solid Tumor Fusion Panel assays:
            1. Illumina products: the vendor notifies the lab of any available updates/upgrades to reagents, instrumentation, and software. The lab will maintain records of these communications.
            2. Pillar and ArcherDX products: the vendor notifies the lab of any available updates/upgrades to reagents and software. In addition, the Director will discuss any pending updates with technical services. The lab will maintain records of these communications.
            3. All additional equipment, including ancillary instruments and software: Lifespan security policies and requirements prevent the automatic update of electronic systems.
         2. Upon any changes to the procedures, performance specifications will be evaluated. If appropriate, revalidation/confirmation should be performed.
   2. Procedure for monitoring, recording, and implementing patch releases, upgrades, and other updates to the bioinformatics pipeline.
      1. Assays have been validated using specific versions of bioinformatics/software. Therefore, no upgrades shall occur without appropriate review and, if necessary, validation. However, the lab shall periodically monitor for the availability of updates/upgrades.
      2. For the AMP 2 assay:
         1. PiVAT: the vendor notifies the lab of any available updates/upgrades to reagents and software. In addition, the Director will discuss any pending updates with technical services. The lab will maintain records of these communications.
         2. Lifespan AMP Reporter 2
            1. If available, developers will notify the lab of any available updates/upgrades to software.
            2. Approximately every 12 months, the Director of Clinical Bioinformatics or designee will review software versions and available updates/upgrades. The updates will be discussed with Lab Director and approved, as appropriate.
         3. Upon any changes to the pipelines, performance specifications will be evaluated. If appropriate, revalidation/confirmation should be performed.
      3. For the Solid Tumor Fusion Panel assay:
         1. Archer Analysis: the vendor notifies the lab of any available updates/upgrades to reagents and software. In addition, the Director will discuss any pending updates with technical services. The lab will maintain records of these communications.
         2. Lifespan Fusion Reporter:
            1. If available, developers will notify the lab of any available updates/upgrades to software.
            2. Approximately every 12 months, the Bioinformatician or designee will review software versions and available updates/upgrades. The updates will be discussed with Lab Director and approved, as appropriate. The Director of Clinical Bioinformatics maintains a table to aid in documentation of this review process.
         3. Upon any changes to the pipelines, performance specifications will be evaluated. If appropriate, revalidation/confirmation should be performed.
2. **VARIANT INTERPRETATION:**
   1. Classification, interpretation, and reporting of sequence variants.
      1. For the somatic/tumor assays, variants are classified using the following Variant Tiering system:
         1. Tier I Variants (Strong Clinical Significance)
         2. Tier II Variants (Potential Clinical Significance)
         3. Tier III Variants (Uncertain Clinical Significance)
         4. Tier IV Variants (Benign or Likely Benign)
      2. For each clinical sample, the associated clinical report should include interpretation of the variants with clinical implications, as appropriate.
         1. Interpretation (such as determination of pathogenicity or therapeutic targetability) may be based on a variety of factors, including visualization of the sequencing data and review of public databases, the literature, and historical laboratory data.
         2. Examples of relevant data include frequency of the mutation in the tumor type (e.g., as reported in the COSMIC database), gene-specific functional data, the availability of targeted therapy, patient-specific clinical/pathological factors, literature/references, and information from publicly available databases and other bioinformatics resources.
         3. Public databases include but are not limited to: COSMIC (for determination of significance and frequency of variants in cancer), dbSNP (for determination of population frequencies of variants), MyCancerGenome, ClinVar, cBioPortal, clinicaltrials.gov, as well as the Genome Browsers from NCBI, Ensembl, and UCSC.
         4. Variants should be classified according to clinical significance, in accordance with professional organization guidelines.
   2. Laboratory database of variants identified and/or reported:
      1. For the Actionable Mutation Panel 2, the Laboratory Director, Pathologists, or other designated personnel maintain a table of all runs and all samples. The table includes sample information, as well as testing results.
      2. For the Actionable Mutation Panel 2 and Solid Tumor Fusion Panel:
         1. The lab information system or RICMBLAB$ network directory contain draft reports for all samples.
         2. Variants may be interpreted with each clinical run. However, if a particular variant has been previously seen in the lab for the same tumor type, the interpretation may be (at least partially) derived from the previous report draft.
         3. Because these assays are intended to identify somatic mutations in cancer, variants from previously reported assay runs will not be routinely reclassified unless specifically requested by a treating provider.
         4. In certain cases, the Lab Director/Pathologist may determine that the interpretation for a specific variant has changed significantly and warrants communication to a Provider and/or an update to a previously released clinical report.
3. **REPORTING OF INCIDENTAL OR SECONDARY FINDINGS:**
   1. Genetic findings unrelated to the clinical purpose for testing are reported if appropriate and as applicable to the specific assay.
   2. Actionable Mutation Panel 2:
      1. This assay is designed to detect somatic mutations in cancer specimens and cannot definitively identify germline mutations. However, criteria are used to interpret variants as benign and likely benign and, if appropriate, this is reported using Tier IV of the clinical report.
      2. In most cases, synonymous and intronic variants are not reported; however, if clinically relevant, these variants may be reported by the Director/Pathologist with appropriate comment.
   3. Solid Tumor Fusion Panel:
      1. This assay is designed to detect somatic gene fusions in cancer specimens and cannot definitively identify germline mutations. However, criteria are used to interpret variants as benign and likely benign and, if appropriate, this is reported using Tier IV of the clinical report.
4. **BIOINFORMATICS:**
   1. Lifespan AMP Reporter2 (LAR 2): this custom software was developed to analyze Next Generation Sequencing (NGS) data produced by the MiSeqDx instrument. It combines multiple open-source programs and scripts as a pipeline. Please refer to AMP 2 Procedure for details on the procedure and programs.
   2. Lifespan Fusion Reporter: this custom software was developed to analyze Next Generation Sequencing (NGS) data produced by the NextSeq instrument. It combines multiple open-source programs and scripts as a pipeline. Please refer to Solid Tumor Fusion Procedure for details on the procedure and programs.
   3. Lifespan Fusion Review Tool and All\_Fusions Tool: these custom tools were designed to aid in analysis of NGS data. Please refer to Solid Tumor Fusion Procedure for details on the procedure and programs.
   4. The source code for programs included in Bioinformatics Pipelines are maintained on the Lifespan server and backed-up, as appropriate.
      1. A list of the pipeline versions, as well as the versions of any included programs is maintained on the Lifespan network drive.
      2. Actionable Mutation Panel 2
         1. The bioinformatics pipeline was validated, and software validation execution results are maintained in the validation documentation on the Lifespan network drive.
         2. Source codes: LAR includes 7 open sources programs as well as custom Perl and shell scripts to automatically run the pipeline, remove meta-files and move all FASTQ, BAM, VCF and annotated files into Run\_Files folder. The scripts are included in the Actionable Mutation Panel Procedure 2: Appendix A.
      3. Lifespan Fusion Reporter:
         1. The bioinformatics pipeline was validated, and software validation execution results are maintained in the validation documentation on the Lifespan network drive.
         2. Source codes: LFR includes 4 open-source programs, as well as custom Perl scripts that create shell scripts to run the programs and perform upstream analysis. The scripts are included in the Solid Tumor Fusion Panel Procedure: Appendix A.
      4. Any version updates will be documented and maintained on the Lifespan network drive.
      5. Upgrades in the pipeline and revalidation: the Director of Clinical Bioinformatics will manually do upgrades in the programs (if necessary) and revalidate the new version of the programs using samples previously tested. Revalidation data will include comparison of the variants detected by the previous and new versions of the program.
   5. Sequence read alignment to a reference sequence:
      1. The reference sequence assembly, version number, source, and URL from where the reference assembly was downloaded shall be recorded, as well as details of any modifications made to the reference file.
      2. LAR 2 includes several human reference files:
         1. human\_g1k\_v37.fasta: This reference file is used in BWA for raw read alignment. No modifications were made on this reference file. The file can be downloaded from NCBI link below:
            1. ftp://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/technical/reference/human\_g1k\_v37.fasta.gz

PLEASE NOTE that the above links are working as of the date of LAR development; the links may not be available in the future, as the reference files get updated by the source.

* + 1. Reference Files for Lifespan Fusion Reporter:
       1. human\_v90: This reference file is used in FusionCatcher for raw read alignment. It can be downloaded from the link below:
          1. <https://sourceforge.net/projects/fusioncatcher/files/data/>
       2. GRCh37\_v19\_CTAT\_lib\_Feb092018: This reference file is used in STAR-Fusion for raw read alignment and fusion detection. It can be downloaded from the link below:
          1. https://data.broadinstitute.org/Trinity/CTAT\_RESOURCE\_LIB/\_\_genome\_libs\_StarFv1.3/
  1. Description of input and output data files or information in each process step
     1. LAR2: a diagram with accompanying description can be seen in the Actionable Mutation Panel 2 Procedure.
     2. Lifespan Fusion Reporter: a diagram with accompanying description can be seen in the Solid Tumor Fusion Panel Procedure.
  2. Criteria and specific thresholds used for inclusion of variants in classification and interpretation steps following variant calling:
     1. Actionable Mutation Panel 2:
        1. For both indels and base substitutions (single and multi-nucleotide variants), the variant call thresholds are:
           1. Total depth of coverage ≥200x
           2. Variant allele fraction ≥5%
     2. Solid Tumor Fusion Panel:
        1. The fusion call thresholds in the Archer Analysis software are:
           1. The number of unique start sites supporting the event ≥5
           2. The number of unique reads supporting the event ≥5
  3. Bioinformatics processes and thresholds applied for prioritizing and identifying causal or candidate variants or genes:
     1. Public databases are reviewed for classification of pathogenicity, as described above.
  4. Acceptance and rejection criteria for the results generated by the analytical bioinformatics process are described above, according to the run metrics and sample metrics.
  5. When results fail to meet the laboratory's acceptance criteria, refer to the assay procedure for appropriate next steps and/or corrective actions.
  6. Infrastructure for NGS Analytical Bioinformatics Pipelines:
     1. The analytical bioinformatics processes and pipelines for NGS are operated within the larger Lifespan IT infrastructure. The infrastructure includes the hardware, software, networks, data centers, facilities, and related equipment used to operate and/or support IT services.
        1. The IT infrastructure of the bioinformatics pipeline and computer environment must be tested during validation and continuously monitored during clinical testing. This will be performed through the monitoring of quality control metrics for each assay, as well as CAP proficiency testing to ensuring accuracy of results.
        2. Test validations should be performed within the environment in which the pipeline is expected to be executed in clinical production.
     2. lsmplinux3:
        1. Description of the IT infrastructure for the Server, lsmplinux3:
           1. The bioinformatics pipeline for Actionable Mutation Panel 2 is executed on a server, lsmplinux1 with Red Hat Enterprise Linux Server release 8.6.
           2. Specifications including hardware, firmware, CPU, memory, and hard disk are included in Next-Generation Sequencing Infrastructure Policies Appendix A.
           3. IP address of the server is 10.37.28.238.
        2. Disaster Recovery Plan:
           1. The server, lsmplinux3 is managed by Lifespan IT. In the event of a disaster involving the server, the Director of Clinical Bioinformatics, Bioinformatics Analyst, or appropriately designated personnel will contact Lifespan IT and submit a HEAT Ticket (https://lifespan.saasit.com/) describing the details of the incident. The files on the server are backed-up nightly by Dell EMC Networker Software. In the event of a disaster, the files will be replaced by Lifespan IT.
        3. Downtime Procedure:
           1. If the server needs to be rebooted or shut down by Lifespan IT for an update or any other reason, the Director of Clinical Bioinformatics is informed in advance. Depending on the duration of the downtime, the time of the shutdown is adjusted according to the clinical NGS run schedule to prevent any interruptions in the bioinformatics analysis. In the event of an unexpected downtime interrupting bioinformatics analysis, the data produced during the interrupted run is deleted and the analysis is restarted from FASTQ files.
        4. Access to the Server:
           1. Access to the server, pipelines, and associated data is granted by Lifespan IT. The Director of Clinical Bioinformatics or Lab Director will make a request to grant access for a user by submitting a HEAT ticket to Lifespan IT. Each user is given their own login information. The programs in the pipelines are assessed for risk and approved by Lifespan IS before they are installed on the server by Lifespan IT. The users are not allowed to install any programs.
        5. Electronic File Transfer:
           1. The file transfer between a computer or a shared directory on Lifespan network to the server is done through WinSCP installed on the computer by Lifespan IT. The file transfer is using SFTP protocol. The login to WinSCP is password-protected and every user uses their own login information to transfer data.
           2. The checksum is calculated on the FASTQ files stored on the MiSeq, lsmplinux3, and RICMBLAB$ network drive. The values are compared in an Excel file to ensure that data integrity is preserved pre- and post-transfer of data. The Excel file is saved in the AMP 2 run folder on the RICMBLAB$ shared drive within the subfolder containing the FASTQ files.
        6. Information Available from Lifespan IT upon request:
           1. Records containing the level of service (e.g., uptime, monitoring, technical support availability) and redundancy (i.e., for data storage and computers) provided by the IT infrastructure.
           2. Records of testing the disaster recovery procedure.
     3. lsmplinux2:
        1. Description of the IT infrastructure for the Server, lsmplinux2:
           1. The bioinformatics pipeline for Solid Tumor Fusion Panel is executed on a server, lsmplinux2 with Red Hat Enterprise Linux Server release 7.9.
           2. Specifications, including hardware, firmware, CPU, memory, and hard disk, are included in the Next-Generation Sequencing Infrastructure Policies Appendix B.
           3. IP address of the server is 10.217.75.25.
        2. Disaster Recovery Plan:
           1. The server, lsmplinux2 is managed by Lifespan IT. In the event of a disaster involving the server, the Director of Clinical Bioinformatics, Bioinformatics Analyst, or appropriately designated personnel will contact Lifespan IT and submit a HEAT Ticket (<https://lifespan>.saasit.com/) describing the details of the incident. The files on the server are backed-up nightly by Dell EMC Networker Software. In the event of a disaster, the files will be replaced by Lifespan IT.
        3. Downtime Procedure:
           1. If the server needs to be rebooted or shut down by Lifespan IT for an update or any other reason, the Director of Clinical Bioinformatics is informed in advance. Depending on the duration of the downtime, the time of the shutdown is adjusted according to the clinical NGS run schedule to prevent any interruptions in the bioinformatics analysis. In the event of an unexpected downtime interrupting bioinformatics analysis, the data produced during the interrupted run is deleted and the analysis is restarted from FASTQ files.
        4. Access to the Server:
           1. Access to the server, pipelines, and associated data is granted by Lifespan IT. The Director of Clinical Bioinformatics or Lab Director will make a request to grant access for a user by submitting a HEAT ticket to Lifespan IT. Each user is given their own login information. The programs in the pipelines are assessed for risk and approved by Lifespan IS before they are installed on the server by Lifespan IT. The users are not allowed to install any programs.
        5. Electronic File Transfer:
           1. The file transfer between a computer or a shared directory on Lifespan network to the server is done through WinSCP installed on the computer by Lifespan IT. The file transfer is using SFTP protocol. The login to WinSCP is password-protected and every user uses their own login information to transfer data.
           2. The checksum is calculated on the FASTQ files stored on the NextSeq, lsmplinux2, and RICMBLAB$ network drive. The values are compared in an Excel file to ensure that data integrity is preserved pre- and post-transfer of data. The Excel file is saved in the Fusion run folder on the RICMBLAB$ shared drive within the subfolder containing the FASTQ files.
        6. Information Available from Lifespan IT upon request:
           1. Records containing the level of service (e.g., uptime, monitoring, technical support availability) and redundancy (i.e., for data storage and computers) provided by the IT infrastructure.
           2. Records of testing the disaster recovery procedure.
     4. Desktop server with Archer Analysis Program.
        1. This computer is located in the Lifespan Data Center and is a desktop server maintained by Lifespan IT.
        2. The Archer Analysis program is executed on the server, lsarcher02 with ESXi 6.5 (VM version 13).
        3. IP address for the server with the Archer Analysis Program is 10.37.128.27.
        4. All Lifespan IT policies for disaster recovery, downtime, access, security, and IP address assignment apply to this server. Data recovery requires a complete re-format of the system. Steps for performing a System Rescue can be seen in the Next-Generation Sequencing Infrastructure Policies Appendix C.

1. **ATTACHMENTS:**
   1. Next-Generation Sequencing Infrastructure Policies Appendix A
   2. Next-Generation Sequencing Infrastructure Policies Appendix B
   3. Next-Generation Sequencing Infrastructure Policies Appendix C
2. **REVISIONS:**
   1. 1/15/2020: Updated footer with new laboratory name, added section for Infrastructure of Bioinformatics Pipeline.
   2. 12/14/2020: Additional steps included for Solid Tumor Fusion Panel.
   3. 7/6/2021: Updated information on File Transfer and the checksum calculation.
   4. 11/18/2022: Updated information on data management and Archer Analysis Program IP address.
   5. 1/16/2024: Updated information on Actionable Mutation Panel 2, Pillar Variant analysis Toolkit and Lifespan AMP Reporter 2. Added lsmplinux3 and lsmplinux2 infrastructure information to Appendix A and B, respectively.