LIMS Integration of the Agilent 4200 TapeStation System

Technical Overview

Authors

Florian Jostock, Laura-Jane Behl, and Stephan Wolf
DKFZ Genomics and Proteomics Core Facility,
High Throughput Sequencing Unit,
Heidelberg, Germany

Elisa Viering
Agilent Technologies, Inc.
Waldbronn, Germany

Abstract

At the Genomics and Proteomics Core Facility in the German Cancer Research Center (DKFZ), the Agilent 4200 TapeStation System is an established quality control tool for analysis of incoming samples and intermediates during the automated library preparation workflows for high-throughput sequencing. This Technical Overview describes the integration of the 4200 TapeStation data in the laboratory information management system (LIMS).

Agilent Technologies

Introduction

The German Cancer Research Center (DKFZ) is one of the largest biomedical research institutions in Germany. The High Throughput Sequencing Unit of the DKFZ Genomics and Proteomics Core Facility provides sequencing services to national cancer genome projects for multiple applications.

It is essential to know the quality of incoming samples for the experimental success of sequencing. The High Throughput Sequencing Unit subjects received DNA or RNA samples to an incoming quality control (QC) analysis, and performs additional QC steps with the Agilent 4200 TapeStation System before sequencing[1,2]. The laboratory information management system (LIMS) is implemented to manage sample and sequencing data. This Technical Note describes how the DKFZ integrated the 4200 TapeStation data into their own LIMS system.

Technical Details

The 4200 TapeStation System (G2991AA) with the Agilent 4200 TapeStation Software version A.02.01 (SR1) and the ScreenTape devices and reagents were obtained from Agilent Technologies (Waldbronn, Germany). The Barcode scanner was provided by (Conrad Electronics SE, Model: LS6300E)

The described LIMS was programmed by the DKFZ using Java Enterprise Edition with JavaServer Faces (JSF) technology for templating and PrimeFaces as JSF library. The LIMS stores its data in a Microsoft SQL Server database. The database abstraction layer MyBatis ensures the LIMS stays database-independent, and could be moved to another database in the future. As an application server the open-source web server Apache Tomcat was used.

Results and Discussion

LIMS

When analyzing RNA or DNA samples, using the 4200 TapeStation System, two slightly different approaches (tube and plate workflow) are applied, depending on the number of samples (Figure 1).

The customer submits up to 80 samples in one batch (Submission) using the LIMS, and provides information such as sample name and concentration. The sample name is provided in a free text field (Figure 2).

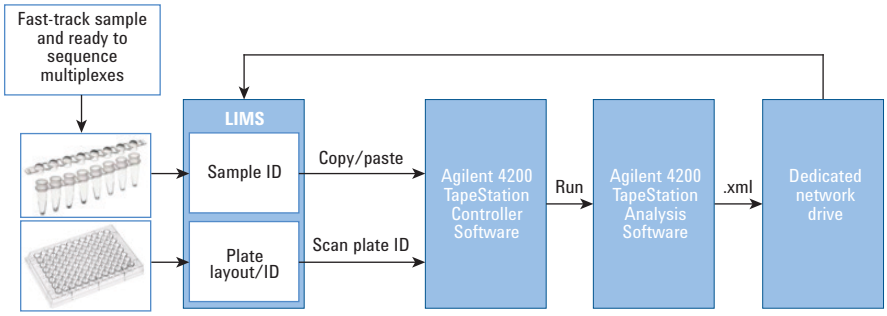

Figure 1. LIMS integration of the Agilent 4200 TapeStation System at the DKFZ.

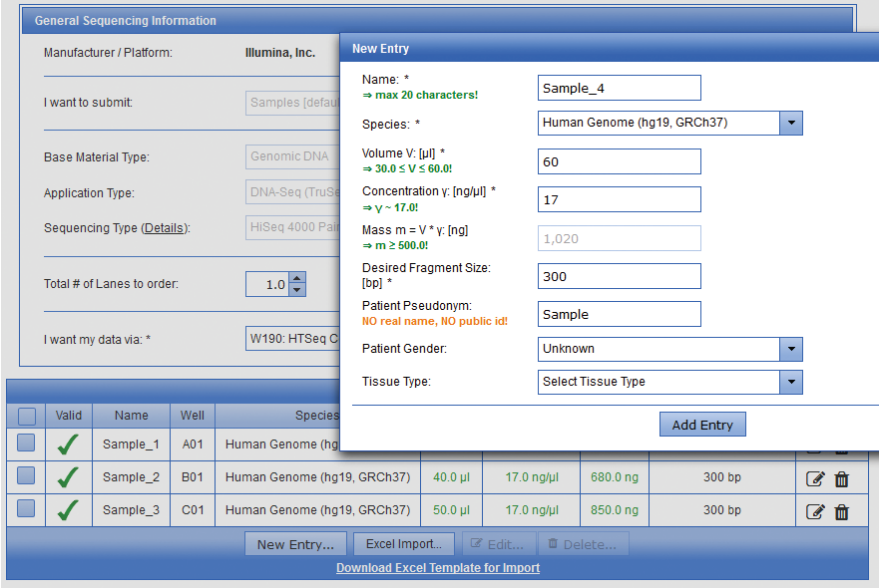

Figure 2. Submission form: the user enters sample information such as sample name and concentration.

The LIMS generates a unique number (ASID) for each sample, automatically arranges the submitted samples in a plate layout, and creates a unique plate ID. The customer provides the samples in plates. Exceptions are specific fast-track samples that require immediate analysis and ready-to-sequence multiplexes.

Agilent 4200 TapeStation Controller Software/tube workflow

For each TapeStation run, the LIMS creates a sample ID. The submission number and ASID are assigned with the sample name from the free text field to create the sample ID (submission number_ASID_sample name). The sample name is part of the sample ID to enable customers to easily identify their samples in the final report they will receive. The only requirements for the LIMS integration are submission number and ASID.

The sample IDs can be copied from the LIMS System (Figure 3, named Label BA/TS in the LIMS), and pasted into the 4200 TapeStation Controller Software (Figure 1 Tube workflow).

Agilent 4200 TapeStation Controller Software/plate workflow

The number of samples per submission varies. To enable high-throughput capability, different submissions are arranged on one plate, such as for automated library preparation. Therefore, a plate layout is created in the LIMS, then the LIMS generates a plate ID (Figure 3B).

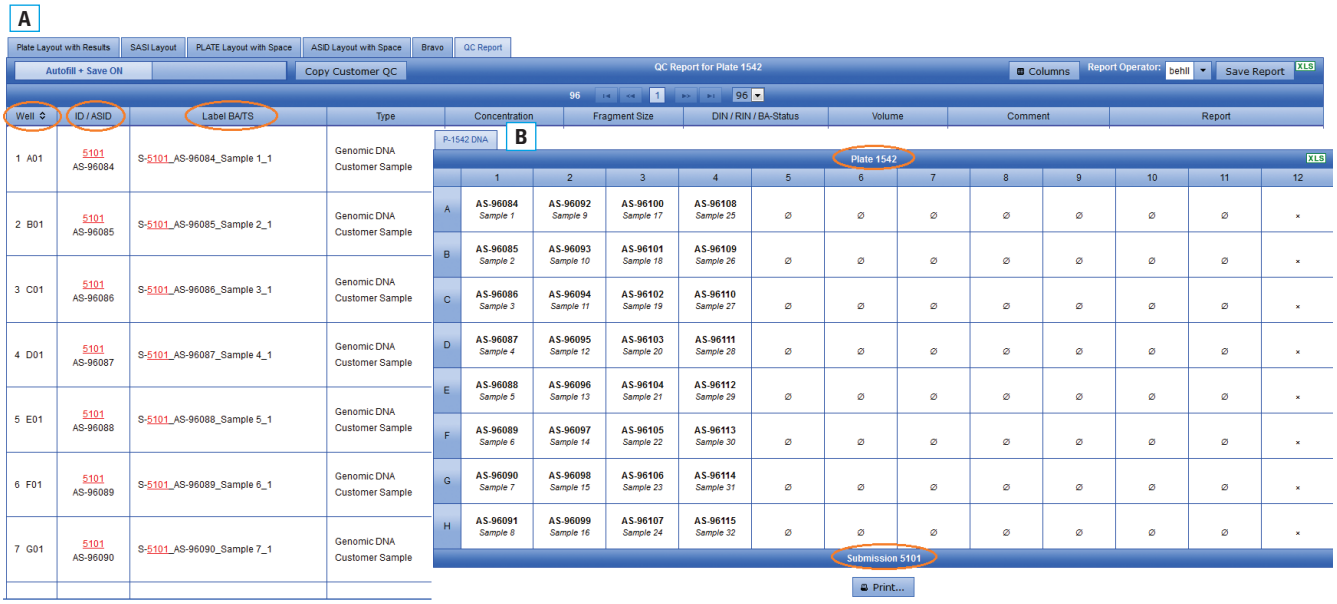

Figure 3. Sample view in the LIMS. A) Table view including the well position submission ID, ASID, and sample ID (Label BA/TS). B) Plate layout view; on top is the plate ID, on the bottom is the submission ID. At the well positions is the ASID and the sample name.

The plate ID is scanned into the User field of the 4200 TapeStation Controller Software with a handheld barcode scanner. The sample identification is based on the plate ID and well location (Figure 1 plate workflow and Figure 4). Additionally, sample IDs can be inserted through copy and paste. In this case, the LIMS will double-check if the ASID corresponds to the appropriate submission number and well position in the plate.

The subsequent workflow remains the same, independent of whether sample identification is provided by sample ID or plate ID. The lot numbers of the Tapes and Regents are scanned into the Notes field (Figure 4). Next, the 4200 TapeStation Controller Software starts the run.

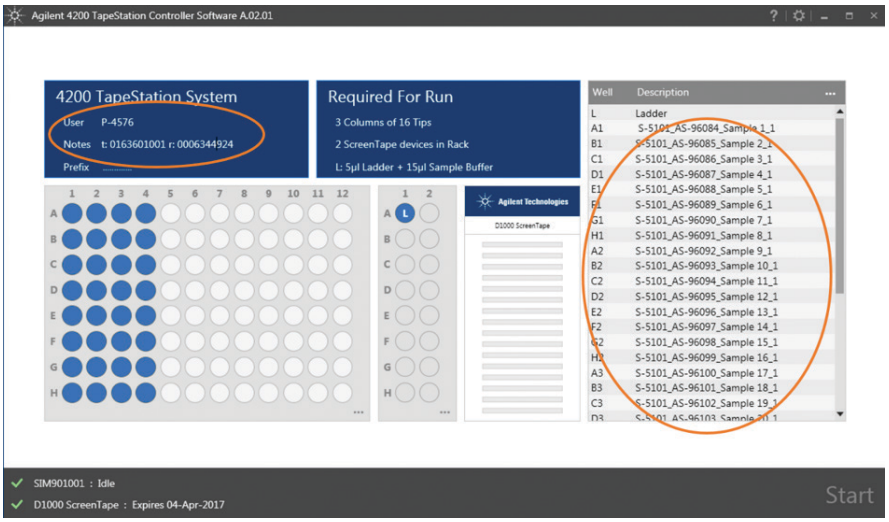

Figure 4. Agilent 4200 TapeStation Controller Software with scanned plate ID and reagent lot numbers and copied sample IDs.

Agilent 4200 TapeStation Analysis Software

Once all samples are processed with the 4200 TapeStation System, the 4200 TapeStation Analysis Software automatically opens, and the user can review and process the results. A report file (.doc or .pdf format) is created. The data are then manually exported as an .xml file and saved on a dedicated network drive. A dedicated network drive with folders for the individual instruments must be selected as the storage location within the 4200 TapeStation Analysis Software using the browse function (Figure 5). The data-dedicated network drive is used as a storage repository for the 4200 TapeStation output files. The LIMS scans this folder, and imports new data.

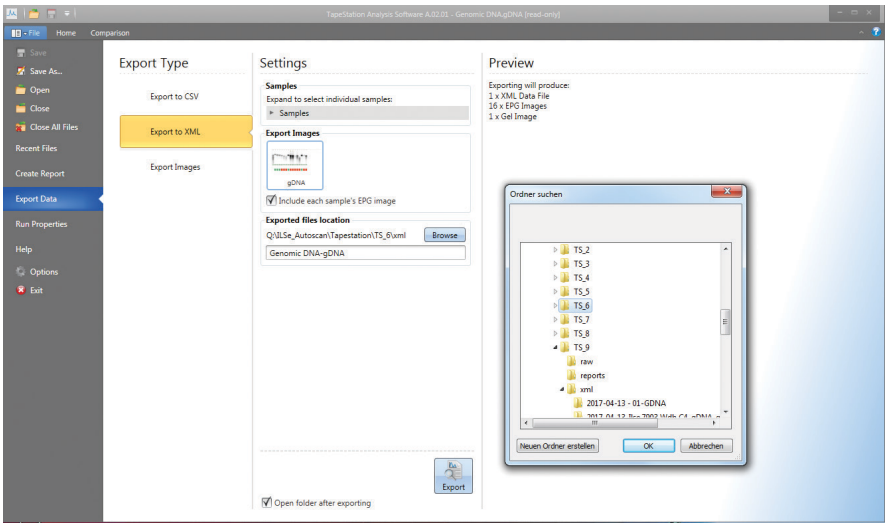

Figure 5. Agilent 4200 TapeStation Analysis Software shows how to create .xml file, and the selection of the storage location in the dedicated network drive.

.XML File

The first part of the .xml file contains general information such as file name, storage location, the assay used, date and time of the run, and notes (scanned lot numbers). In addition, it includes detailed information on the ScreenTape device used and the system environment. System environment includes the scanned plate ID, under <Experimenter>, and information such as instrument type and software version (Figure 6). It also comprises the units for concentration, molarity, and size of the sample.

In addition, the .xml file contains all sample results including well number, sample name (copied sample ID), concentration, and, depending on the assay, the DNA integrity number (DIN) for gDNA, or the RNA Integrity Number (RIN^e) for RNA, as shown in Figure 7.

```
<?xml version="1.0" encoding="utf-8"?>
<File>
 <FileInformation>
  <FileName>C:\Users\eviering\Documents\Agilent\TapeStation Data\Demo Files\
            \gDNA-Tubes-16.GDNA</FileName>
  <Assay>Genomic DNA</Assay>
  <Comment>
  </Comment>
  <Study>
  </Study>
  <RunEndDate>24-Jul-2015 1:22 PM</RunEndDate>
  <DINVersion>2.2.7.9032</DINVersion>
 </FileInformation>
 <ScreenTapes>
  <Notes>
  <T:0163601001 R:0006344924/Notes>
  <ScreenTape>
   <ScreenTapeID>01-S025-150625-01-000014</ScreenTapeID>
   <ExpiryDate>07-Aug-2015 12:58 PM</ExpiryDate>
   <ScreenTapeHistory>First run 24-Jul-2015, 1 run(s) performed</ScreenTapeHistory>
   <ElectrophoresisTemp>26.2</ElectrophoresisTemp>
   <ElectrophoresisTime>210</ElectrophoresisTime>
   <TapeRunDate>24-Jul-2015 12:59 PM</TapeRunDate>
   <Environment>
    <Experimenter>P-4576</Experimenter>
    <InstrumentType>32768</InstrumentType>
    <InstrumentSerialNumber>DEDAA00054</InstrumentSerialNumber>
    <Computer>TapeStationLaptop</Computer>
    <ControllerVersion>2.1.18.8137</ControllerVersion>
    <OperatingSystem>Microsoft Windows NT 6.1.7600.0</OperatingSystem>
    <AnalysisVersion>2.1.18.8137</AnalysisVersion>
   </Environment>
  </ScreenTape>
 </ScreenTapes>
 <Assay>
  <Units>
   <ConcentrationUnit>ng/µl</ConcentrationUnit>
   <MolarityUnit>nmol/l</MolarityUnit>
   <MolecularWeightUnit>bp</MolecularWeightUnit>
  </Units>
 </Assay>
```

Figure 6. First part of the .xml file generated by the Agilent 4200 TapeStation Software, containing general information including file name, storage location, run time, as well as details on the ScreenTape device and system environment. Information that was scanned in the 4200 TapeStation Controller Software are marked green.

```
  <Samples>
<Sample>
 <WellNumber>A1</WellNumber>
 <Comment>S-5101_AS-96084_Sample1_1</Comment>
 <Concentration>73.4</Concentration>
 <Observations>Ladder</Observations>
 <Alert>
 </Alert>
 <RNA />
 <DIN>8</DIN>
 <ScreenTapeID>01-S025-150625-01-000014</ScreenTapeID>
```

Figure 7. Sample-specific information included in the .xml file generated by the Agilent 4200 TapeStation Analysis Software, for example, well number, sample name, concentration, and DIN or RIN^e. Information (sample ID) that was copied in the 4200 TapeStation Controller Software is marked green.

For each sample, peak information is included in the .xml file generated by the 4200 TapeStation Software. The data are arranged by ascending peak size. Peak area, peak height, peak molarity, and observations (Figure 8) are listed for each peak, including markers of the assay.

Region information is also included in the .xml file, if this feature was used in the 4200 TapeStation Analysis Software.

```
<Peak>
    <Area>3.66</Area>
    <AssignedQuantity>-</AssignedQuantity>
    <CalibratedQuantity>31.1</CalibratedQuantity>
    <Comment>
    </Comment>
    <FromMW>18639</FromMW>
    <FromPercent>38.4</FromPercent>
    <Height>778.229</Height>
    <Molarity>-</Molarity>
    <Size>&gt;60000</Size>
    <Number>1</Number>
    <Observations>
    </Observations>
    <PercentIntegratedArea>89.53</PercentIntegratedArea>
    <PercentOfTotal>74.08</PercentOfTotal>
    <RunDistance>42.5</RunDistance>
    <ToMW>&gt;60000</ToMW>
    <ToPercent>46.3</ToPercent>
</Peak>
```

Figure 8. Example for peak information included in the .xml file generated by the Agilent 4200 TapeStation Software.

Using the import functionality of the LIMS permits automatic import of all new data files from the network drive to the LIMS (Figure 9). The sample assignment is performed based on the unique sample ID, or the plate ID and well location.

Previously defined data from the .xml file (Table 1) are integrated into the LIMS. In addition, the electropherograms and gel images (.png files) are imported.

For the successful LIMS integration of the 4200 TapeStation System, the different units used for the concentration (Table 1) needed to be considered because concentration and unit are read out separately. The high-sensitivity DNA and RNA assays use a different concentration unit (pg/µL) than the other assays (Table 1). The concentration unit of each assay can be found following the .xml path: **/File/Assay/Units/ConcentrationUnit** (Figure 6). The concentration in pg/µL, determined by the high-sensitivity assays of the 4200 TapeStation System, will be converted into ng/µL by the LIMS.

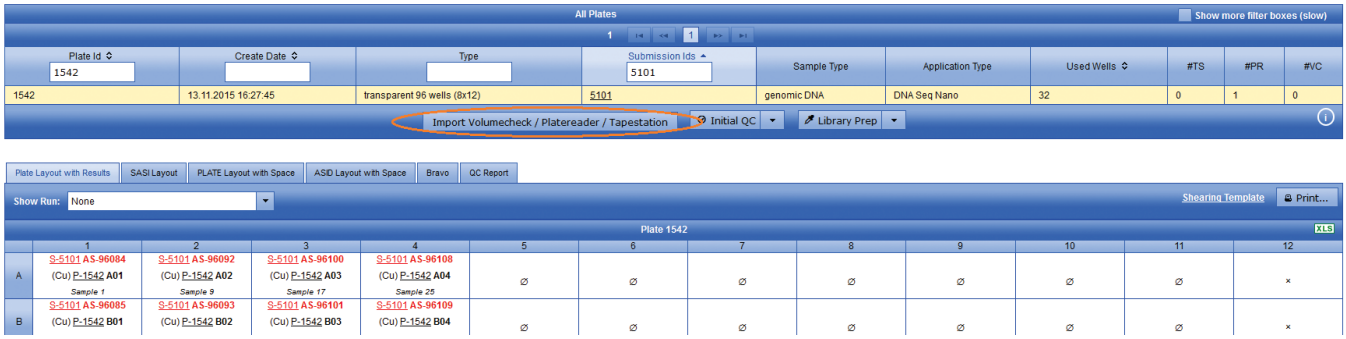

Figure 9. Import function in the LIMS.

Table 1. Overview of ScreenTape assays and the correlating type of data imported into LIMS.

Level	Data	gDNA	D1000	HS D1000	RNA	HS RNA
Sample	Name	X	X	X	X	X
	Well number	X	X	X	X	X
	Comment	X	X	X	X	X
	Concentration	X[1]	X[1]	X[2]	X[1]	X[2]
	DIN	X				
	RINe				X	X
Region	Average size (bp)	X	X	X		
	Region concentration	X[1]	X[1]	X[2]		
Peak	Area	X	X	X	X	X
	Height	X	X	X	X	X
	Molarity (nmol/L)	X	X	X	X	X
	Size	X[3]	X[3]	X[3]	X[4]	X[4]
	Observations	X	X	X	X	X

[1] In ng/µL
[2] In pg/µL
[3] In bp
[4] In nt

The individual electropherograms for each sample, exported as .png files, are assigned using the name of the .xml file and the well location, for example, Genomic DNA_B1. Correct assignment of the gel images based on the .xml file name and the assay was achieved using the assay name from the .xml file (**/File/FileInformation/Assay**). The following modifications were performed to generate the gel suffix:

- The spaces at the beginning and end were removed.
- *High sensitivity* at the beginning or end was removed and replaced with *HS* at the beginning.
- *Genomic DNA* was changed to *gDNA*.

This was mandatory, as the syntax of the assay types differ between the .xml and file name with respect to upper and lower case and spaces. With this approach, it was possible to relate the sample information and the gel image in the LIMS.

Conclusion

This Technical Overview describes the successful LIMS integration of the Agilent 4200 TapeStation System as a QC tool for the sequencing workflow at the Genomics and Proteomics Core Facility of the DKFZ. The benefit of using the 4200 TapeStation System as a QC tool is full automation, and the reduction of hands-on time. In addition, LIMS integration of the 4200 TapeStation System further reduces manual data entry steps, resulting in increased efficiency, and elimination of a potential error source.

References

1. Evaluating the Agilent 4200 TapeStation System for High Throughput Sequencing Quality Control, *Agilent Technologies*, publication number 5991-6892EN, **2016**.
2. Use of the Agilent 4200 TapeStation System for Sample Quality Control in the Whole Exome Sequencing workflow at the German Cancer Research Center (DKFZ), *Agilent Technologies*, publication number 5991-7615EN, **2016**.

www.agilent.com/chem

Published in the USA, May 17, 2017
5991-7984EN

Agilent Technologies