Genome Scan

Next Generation Sequencing

For Illumina HiSeq and NextSeq Technology

Guidelines

genomescan.nl

GenomeScan's Guidelines for Successful NGS Experiments

Using Illumina's HiSeq and NextSeq Technology

Dear customer,

As of the beginning of 2015 ServiceXS became a trademark of GenomeScan B.V. GenomeScan focuses exclusively on Molecular Diagnostics whereas our ServiceXS trademark is intended for your R&D projects.

GenomeScan is dedicated to help you design and perform Next Generation Sequencing (NGS) experiments. This guide provides information and resources enabling you to derive optimal results from your project when using the Illumina NGS platforms at GenomeScan.

Read this guide carefully! This document will inform you about the various stages in our service process, and enables us to tune mutual expectations. Please note that your quotation can deviate from the general conditions presented in this guideline.

GenomeScan has extensive experience in performing sequencing experiments starting from experiment design to data analysis. We can advise and assist you in every step of your project.

Furthermore, we provide basic, standard or custom bioinformatics solutions. Since NGS experiments result in vast amounts of data this can be quite challenging. Our ability to assist in the analysis of results can be the key factor leading to a successful project.

Do not hesitate to contact us if you have any questions after reading this guideline!

On behalf of the GenomeScan team,

Project Manager NGS	David van der Meer
Project Manager Diagnostics	Stephanie van den Oever
Project Manager Bio-IT	Thomas Chin-A-Woeng
Laboratory Coordinator	Floor Pepers
Laboratory Manager	Mark de Jong
Technicians	Gerard Gathier, Remco Stolk, Gaby Rijsdijk, Niels de Water, Fahim Behrouz

TESTING
RvA L 518

Document Outline

Changes to Previous Version (5.0)
-Minor changes
-Added Rapid Mode as option for the HiSeq 2500
-Added NEBNext Ultra Directional RNA Library Prep Kit for Illumina
-Added NEBNext Small RNA Library Prep Kit for Illumina
-Added NEBNext dual index kit
-Table 3 updated with new expected yields and sequencing options

Changes to Previous Version (5.1)
-Minor changes
-Discontinuation of TruSeq RNA Sample Preparation Kit v2
-Implementation of NEBNext Ultra Directional RNA Library Prep Kit for Illumina

Changes to Previous Version (5.2)
-Added NextSeq500 as a validated sequencing platform.
-Added HiSeq2500 Rapid Mode as a validated sequencing option
-Added Fragment Analyzer as the standard method for quality assessment of the DNA and RNA samples
-Added Success rate policy
-Amendment on the Spare and Replacement sample policy
-Accreditation logo clarified. Now connected to GenomeScan B.V.
-Lay-out changes

Changes to Previous Version (6.0)
-Added HiseqSeq4000 as a validated sequencing platform
-Added WGBS and RRBS as Methylseq options
-Added NEBNext Microbiome DNA Enrichment Kit
-Added NEXTflex 16S V4 Amplicon-Seq Kit 2.0 for Illumina
-Added NuGen Ovation Ultra Low Methylseq Library Systems
-Added NuGen Ovation RRBS Methyl-Seq System
-Updated sample requirements
-Minor textual changes

ServiceXS™ | GenomeScan B.V. | Plesmanlaan 1D | 2333 BZ Leiden | The Netherlands
Telephone: +31 (0)71 568 1050 | Fax: +31 (0)71 568 1055
info@genomescan.nl | info@servicexs.com | www.genomescan.nl | www.servicexs.com

Chapter 1 Service description

1.1 Introduction

Next Generation Sequencing (NGS) is a cutting edge technology that enables you to take a whole new approach in answering your research questions. Illumina's NGS platform uses Solexa Sequencing technology, which allows massive parallel sequencing of millions of fragments by using reversible terminator-based sequencing chemistry. Each run results in gigabases of high quality data in a minimal amount of time, at reduced costs compared to conventional sequencing methods.

When using our NGS Services, you have access to this high-throughput sequencing technology, our knowledge and expertise, without having to invest in expensive equipment, qualified personnel and optimization of the required methods. Not all customers have the same background with regards to performing NGS experiments and subsequent data analyses. So we can help you at various stages of your project.

1.2 Your project at GenomeScan

Successful NGS studies start before the DNA and RNA samples are submitted to GenomeScan: experimental design, sample preparation, sample purity and concentration assessment all contribute to the overall success. To provide you from high quality data GenomeScan has a strict policy concerning sample preparation and sample shipment and their timelines (see chapter 7).

GenomeScan offers three different type of NGS services using the Illumina platform options:

- **Standard Service:** GenomeScan performs all steps to generate raw data files. A comprehensive overview of all specific parts of the sequencing workflow is described in Fig. 2. You are regularly informed on the progress of your experiments at predefined time points (see Fig. 1).
  After finalization of the experiment(s) you will receive a project report, data quality report, and your raw data. These reports provide a summary of the experimental procedures that have been performed and give an overview of the Quality Control (QC) measurements of the resulting data. This service does not include secondary data analysis.

- **Ready-to-Run Service:** GenomeScan offers this service to a select group of customers who have in-house experience with NGS and library preparation. GenomeScan performs an accurate concentration measurement on Ready-to-Run samples to optimize cluster generation for sequencing. From this point the workflow is identical to our Standard Service, ending with the same high standard quality check on the resulting data.

- **Data Analysis Service:** Customers can either request a quotation for the data analysis separately, or combine this additional service with Standard or Ready-to-Run Services. Whether or not Data Analysis Service is included in your project, is clearly stated in the quotation. For information about data analysis options, please contact your Project Manager or Sales Representative. Separate guideline documents are available describing in detail the data analysis workflows, deliverables, and terms and conditions.

GenomeScan works according to ISO/IEC 17025 regulations

ISO 17025 is an international standard that ensures that analytical laboratories maintain a high level of technical competency leading to a continuous output of high quality data. The Quality Management System (QMS) implemented at GenomeScan has an impact on all experiments performed as well as the way we communicate with customers and how we deal with our providers. Each experiment is designed to include QCs at pre-set intervals, which ensures correctness and reliability of experiments. The accreditation of GenomeScan is a formal recognition that GenomeScan is a competent partner which expands your research options with cutting edge technologies. Your quotation states if your project will be executed under ISO 17025 regulations.

1.3 General Workflow and Turnaround Time

The turnaround time of your project is dependent on several aspects like *e.g.* application, number of samples, number of reads and desired coverage. Your quotation often contains an expected turnaround time but we will discuss this with you to tune mutual expectations if required. When you sign the quotation you will also fill out the expected sample delivery date on the Purchase Order (PO) form. Upon receipt of the completed PO form, our Project Manager will contact you to discuss further steps, such as sample delivery and project planning. You will also receive a Sample Submission Form, to be filled out to accompany the sample shipment. We will adjust our planning according to the discussed date. Therefore, we strongly advise you to contact your Project Manager in case the samples are delayed and we must postpone your experiment(s). If your samples have not arrived in time, our turnaround time will be affected since the project must be entirely rescheduled.

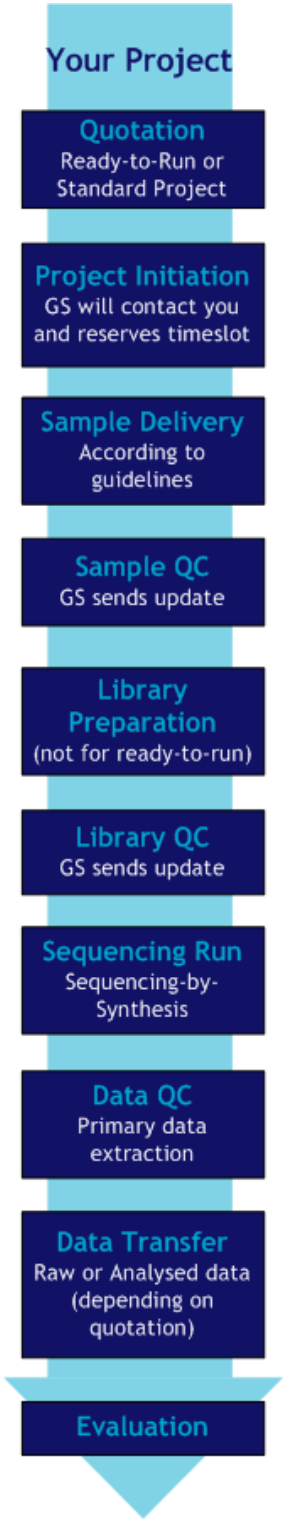
Furthermore, other experimental factors like poor sample quality or low sample concentration might affect the workflow in the lab, which can lead to an increased turnaround time. Please note that for new applications such as Proof-of-Principle (PoP) projects and projects that contain special requirements in general, no time indication can be given to fulfil the requirements specified in the quotation.

After completion of your project, your Project Manager will discuss the objective(s) of the data analysis with you. He/she will also give an indication when the data analysis project is started and when you can expect to receive the data analysis report. Custom data analysis on average takes an additional 2-4 weeks counting from the start of the bioinformatics project, depending on the nature of the analysis and the number of concurrent projects at that time.

The projects that are performed at GenomeScan all share the same workflow that is anchored in our Quality Management System. In Fig. 1 you can find all individual steps that together build up your NGS project. You can also find the intervals at which you receive project updates by email. During the workflow the samples undergo extensive Quality Control at different stages (see Fig. 1). At these time points the project can be halted in case of poor sample quality and the project can be reinitiated with replacement samples if required.

Your project is initiated when the Purschase Order (PO) form is completed, signed and returned to GenomeScan. Our administrator will send you a confirmation email containing the project reference number.

Your Project Manager (PM) will contact you as soon as possible to discuss the project details such as sample delivery, project planning etc. You will receive a Sample Submission Form in which you can fill in the specifications of your samples.



When you have informed GenomeScan of the sample arrival date, the PM makes a reservation in the laboratory planning for your project and the required materials will be ordered.

After arrival of the samples at GenomeScan, quality of the starting material will be assessed, according to Chapter 4. A report with sample QC results is sent to you by email.

When all samples pass sample QC, GenomeScan initiates the sample preparation, followed by QC experiments to assess the quality of the resulting library. Once passed, the sample(s) are queued for clustering and sequencing.

The time required to perform the sequencing run depends on the length of the run, whether the run is Single-Read or Paired-End, whether the samples are indexed and on the sequencer used. Required runtimes range from 11 hours (75 bases Single-Read) on the NextSeq 500 up to 1 weeks on the HiSeq 2500 (125 bp Paired-End). A HiSeq4000 150 bp Paired-End run takes ~3.5 days.

GenomeScan assesses the resulting data quality by analyzing quality metrics of the run. Final inspection of the data takes place and the dataset containing the raw data files is generated for both types of projects; Standard as well as Ready-to-Run. This package is sent to you together with the Project and Quality Control Reports (see Chapter 6).

If you have included data analysis in your service package, our Bioinformatics department starts analyzing the data after sequencing. This takes approximately 2-4 weeks depending on the complexity of the project.

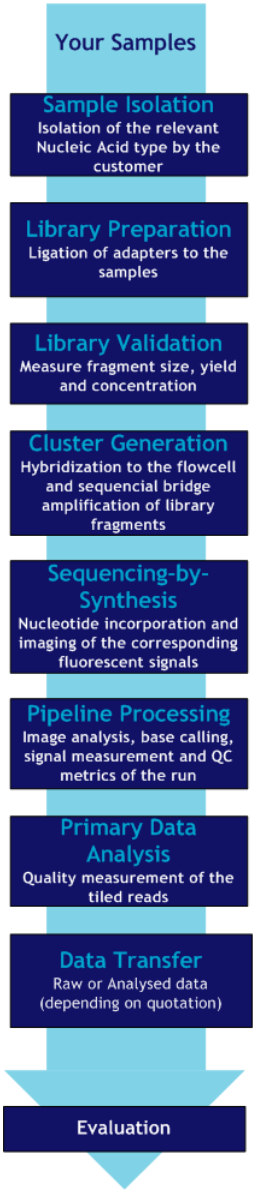
When desired, the customer is invited for a meeting in which your project will be evaluated. The workflow and results of the project are discussed, technical information is provided, and consultation regarding further processing or data analysis of the data is given.

Chapter 2 Experimental Design & Workflow

GenomeScan provides access to many applications ranging from (*de novo*) genome sequencing, targeted sequencing, miRNA, small RNA, ChIP-seq as well as whole transcriptome analysis. Despite the variety in applications GenomeScan provides, all projects are performed using our following procedure.

2.1 Library Preparation

While subsequent steps in the sequencing workflow are more or less standardized across all applications, the sample (or library) preparation is unique and the most critical step in the procedure. Adapter sequences are added to sample fragments, which is the essential first step to enable generation of clusters. GenomeScan performs all types of sample preparations according to our Standard Operating Procedures (SOPs) using validated methodology.



GenomeScan always performs Paired-End sample preparation for sequencing, even for Single-Read runs. This enables us to switch to an additional Paired-End run when desired, thus increasing flexibility. Both Single-Read and Paired-End sample preparation protocols lead to identical data quality.

2.2 Library Validation

After sample preparation, GenomeScan checks the resulting libraries. The purity, fragment size, yield and concentration of the resulting sample preparation are determined. If the quality of the library is according to our specifications, we will proceed to cluster generation.

2.3 Cluster Generation

During cluster generation, the library molecules are hybridized onto an Illumina flow cell. The flow cell consists of an activated glass-based substrate, allowing hybridization of samples onto the flow cell. Subsequently, the hybridized molecules are amplified using bridge amplification. This results in a heterogeneous population of clusters, with each individual cluster consisting of many identical copies of the original template molecule.

GenomeScan runs a validated control sample on each flow cell, independent of the selected sequencing application. This PhiX control is a sample-independent method to assess the quality of the run. The resulting error rate based on the identity of the obtained sequences with the PhiX reference sequence, is used to assess the performance of the flow cell or lane. Furthermore, the PhiX error rate is an excellent tool to monitor data quality over time, to ensure constant, accurate and robust output.

2.4 Sequencing-by-Synthesis (SBS)

The sequencing technology used by the Illumina platform uses fluorescent-labeled nucleotides to sequence millions of clusters present on the flow cell surface. These nucleotides, specially designed for reversible termination, allow each cycle of the sequencing reaction to occur simultaneously in the presence of all four nucleotides. This fluorescently labelled terminator is imaged as each dNTP is added on the growing DNA strands on the flow cell and then cleaved to allow incorporation of the next base. The natural competition between all four alternatives leads to high accuracy and low error.

2.5 Sequencing Equipment

At GenomeScan we are able to handle small to large sized projects for a wide range of RNA and DNA applications. The number of cycles (read length) is dependent on the application and sequencer, which currently varies from 50 to 250 bases. Your Project Manager or Sales Representative is able to advise you which read length is suitable for your type of application. If you prefer a non-standard read length, please inform your Project Manager or Sales Representative so different options can be discussed.

- The **HiSeq 4000** platform runs two flow cells of eight lanes simultaneously and can produce up to 1300-1500 Gb (75% ≥ Q30) per run (2 x 150 bp) in approximately 5 days. Up to 5 billion quality-filtered clusters and up to 10 billion Paired-End reads are being generated in a complete run.

- The **HiSeq 2500** platform runs two flow cells of eight lanes simultaneously and can produce up to 900-1000 Gb (80% ≥ Q30) per run (2 x 125 bp) in approximately 1 week using SBS chemistry version 4. Up to 4 billion quality-filtered clusters and up to 8 billion Paired-End reads are being generated in a complete run. In addition, the **HiSeq 2500** can also run in Rapid Run Mode using 2 flow cells of 2 lanes simultaneously producing up to 300 Gb (75% ≥ Q30) per run (2 x 250 bp) in approximately 60 hrs with Rapid SBS chemistry v2. Up to 0.6 billion quality-filtered clusters and up to 1.2 billion Paired-End reads are being generated in a complete run.

- The **NextSeq 500** sequencer is a highly flexible platform designed for speed and smaller projects. The NextSeq 500 runs one flow cell and can produce 32-39 Gb in Mid-Output mode and 100-120 Gb (75% ≥ Q30) in High-Output mode in approximately 26-30 hours (2x150 bp) according to Illumina specifications. Up to 130 million and 400 million quality-filtered clusters and up to 260 million and 800 million Paired-End reads are being generated in a complete run.

2.6 Sequencing Kits

At GenomeScan kits and procedures are thoroughly validated to guarantee optimal results for your sequencing project(s). Currently, the following kits are used for library preparation and sequencing. Kits in *italics* are currently under validation and available as PoP option. When other kits/methods should **be beneficial** for your research please inform your Project Manager or Sales Representative to discuss different options.

<u>Sequencing kits:</u>

- HiSeq SBS Kit and Cluster Kit v4
- HiSeq Rapid SBS and Cluster Kit v2
- NextSeq 500 Mid Output Kit
- NextSeq 500 High Output Kit

Sample Prep kits:

- TruSeq Small RNA Sample Preparation Kit
- NEBNext rRNA Reduction Kit
- NEBNext poly(A) Kit
- NEBNext Ultra Directional RNA Library Prep Kit for Illumina
- NEBNext Ultra II DNA Library Prep Kit for Illumina
- NEBNext Microbiome DNA Enrichment Kit
- NEXTflex 16S V4 Amplicon-Seq Kit 2.0 for Illumina Library Prep
- Agilent SureSelect^{XT} kits
- NuGen Ovation UltraLow Methylseq Library Systems
- NuGen Ovation RRBS Methyl-Seq System
- NEBNext Multiplex 96 Index Kit

Chapter 3 Sequencing Applications

DNA Analysis Services

Almost any type of DNA can be used for sequencing ranging from small PCR fragments, targeted regions, cDNA, BACs, small microbial genomes to metagenomes or (large) unexplored genomes for which no information is available at all. Depending on the size of the genome and the desired coverage, one may opt for pooling of multiple samples into one lane by multiplex sequencing to reduce sequencing costs.

3.1 Single-Read Sequencing

Single-Read sequencing is the simplest way to utilize Illumina's sequencing platform for almost every application. The Single-Read technology is available at 50 bases reads on the HiSeq 2500/4000 and 75 bases on the NextSeq 500.

3.2 Paired-End DNA Sequencing

During Paired-End sequencing both the forward and the reverse template strands of each cluster are sequenced. Therefore, next to the additional sequence information, it gives you information on the physical distance between the two reads in your (g)DNA or RNA of interest and allows for a more accurate mapping. This long range positional information is required for certain applications such as detection of chromosomal rearrangements like insertions, deletions and translocations. Beside more sequence information it also is a more accurate way to (re-)sequence an entire genome or a large candidate region.
The unique Paired-End sequencing protocol allows you to choose the length of the insert (200–700 bp) and sequence either end of the insert, generating high quality alignable sequence data. When choosing Paired-End sequencing, the amount of data doubles compared to Single-Read sequencing.

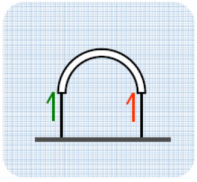


Fig. 3. Paired-End Sequencing
The green arrow indicates the first, forward read, while the red arrow indicates the second, reverse read of the template sequence in a cluster on the flow cell substrate.

NGS generates high quality data to analyze every genome for any species and every genetic variant. Single nucleotide polymorphism (SNP) discovery studies can provide valuable allelic information that can be translated into content for custom genotyping projects. Your sales representative can inform you how GenomeScan aids custom-made projects to generate high throughput SNP data for large cohorts of samples using our unique combination of services.
Sequencing data can also be used for many other applications such as structural variation analysis, gap-closing, copy-number variation (CNV) analysis, or any application in the rapidly expanding portfolio of new applications established by the scientific community.

3.3 Multiplex Sequencing

GenomeScan offers Multiplex Sequencing which allows you to pool samples into a single lane of a flow cell. This increases the number of samples analyzed in a single run, with only minimal increase in the sequencing reagent costs. To prepare samples for multiplexing, a unique coded identifier tag, or index, is added to each sample library. Additional index reads are performed during the run and the

sequencing software pipeline accurately identifies each uniquely-tagged sample for downstream analysis. We routinely use a validated set of 96 of single indices and 96 (dual) indices although many other options are available, depending on the sample preparation method and research questions. Customers may also design their own index tags when preparing their samples in the Ready-to-Run service.

3.4 Agilent SureSelect Target Enrichment

The Agilent SureSelect Target Enrichment system, combined with the BRAVO Automated Liquid Handling Platform, is currently industries' most optimised target enrichment system. With the SureSelect Target Enrichment System, only the genomic areas of interest are sequenced and enable customers to efficiently analyze larger numbers of samples by sequencing just the regions of interest. Combined with the multiplexing capability this extends the benefit by reducing costs and increasing throughput through focussing on selected regions. We have validated the entire workflow starting from genomic DNA which is fragmented and prepared for sequencing, to the end stage in which the targets are captured.

GenomeScan can perform experiments with several off-the-shelf products such as the Agilent Human and Mouse All Exon kits and Inherited Disease Panel. In addition, GenomeScan has ample experience with designing Custom SureSelect kits. If you are interested in using these Agilent based assays, please contact your Sales Representative.

RNA Analysis Services

3.5 Whole Transcriptome Sequencing - mRNA-Seq

Illumina's NGS enables you to profile the transcriptome in any species with mRNA-Seq. Without any prior sequence information, no probes or primers to design, mRNA-Seq delivers unbiased and unparalleled information about the mRNA transcription expression profile of your cells of interest. With RNA sequencing you can generate a full sequence library from any poly-A tailed RNA or rRNA-reduced total RNA sample to characterize all transcriptional activity in coding and non-coding regions, novel transcripts and isoforms, alternative splice sites, rare transcripts, identify regulatory RNAs, and coding region Single Nucleotide Polymorphisms (cSNPs) in one data set.
Our mRNA-Seq protocols are stranded and highly sensitive, allowing detection of common and rare transcripts over six orders of magnitude of dynamic range starting from 5 ng of total RNA. The digital readout by counting the frequency of each sequence in your library, effectively eliminates background noise, traditionally present in micro-array experiments.

3.6 Small RNA Discovery

Small RNA analysis application allows for the discovery and profiling of all known and novel small RNAs of various lengths, coding and non-coding. Using validated procedures RNA libraries can be generated directly from total RNA or small RNA isolations. There is no need for prior sequence or secondary structure information, allowing microRNA discovery and profiling from every species. By multiplexing, the transcriptional activity of various samples can be assessed cost-effective, offering an alternative to micro-array based studies.

Reduced-representation Analysis Service

3.7 ChIP-Sequencing/Complexity Reduction

Chromatin Immunoprecipitation (ChIP) studies allow you to determine the binding sites of your protein of interest to DNA, by crosslinking DNA-protein complexes which are captured using an antibody for a gene of interest. This allows you to study any protein that can be immunoprecipitated from any

living organism, without the need for probe design and optimization. Most transcription binding factor sites can be mapped with relative small generated amounts of sequencing data. This approach delivers low background, a high signal to noise ratio and high specificity at a cost 10 to 30 times less that of microarray-based approaches.
In addition, virtually all techniques, *e.g.* Restriction-site Associated DNA (RAD) sequencing, (custom) sequence capture, complexity reduction of polymorphic sequences or PCR-based enrichments used for complexity reduction of genomes can be sequenced.

3.8 *mtDNA/Microbiome Sequencing*

The NEBNext Microbiome DNA Enrichment Kit employs the MBD2-Fc protein to remove methylated DNA, leaving mainly mtDNA and microbial DNA and other unmethylated regions. If you are interested in sequencing of mtDNA or microbial DNA, this kit may be a suitable option for reduction of gDNA and enrichment of unmethylated regions of interest.

3.9 *16S V4 Amplicon-Seq*

For sequencing of bacterial metagenomes the NEXTflex16S V4 Amplicon-Seq 2.0 Kit is a fast, cost effective, amplicon based option. This kit was designed to prepare multiplexed amplicon libraries that span the fourth hypervariable domain (V4) of microbial 16S ribosomal RNA (rRNA) genes which can be used to identify specific bacterial species.

Methylation Sequencing

3.10 *Whole Genome Bisulfite Sequencing (WGBS)*

Bisulfite sequencing is used for identification of methylated CpGs on a genome wide scale. For WGBS we have validated the NuGEN Ovation UltraLow Methylseq Library kit. With a minimum of 10ng input material a genome wide methylation profile can be created.

3.11 *Reduced Representation Bisulfite Sequencing (RRBS)*

An alternative to Whole Genome Bisulfite Sequencing is Reduced Representation Bisulfite Sequencing. RRBS uses the Msp1 restriction enzyme to select for CCGG motives which are overrepresented in methylated DNA hot spots. As this method enriches for methylated regions, a lower number of sequencing reads are required for sufficient coverage of the selected regions. If you are interested in selected DNA methylation sites, this may be an interesting, cost effective option, for methylation profiling

Chapter 4 Quality Control

Every project starts with a sample QC as part of our standard workflow. Upon receipt of the samples, GenomeScan will inform you that the samples have arrived in good condition. Below, you find an overview of the sample QC criteria in Table 1.

4.1 Sample Quality Control

GenomeScan advises you to perform quality assessments on the tested samples prior to sending them to us. When it is known that high quality samples are provided, this has a positive effect on the workflow. As an extra service, you may deliver 10% surplus samples serving as spare samples. The QC of these spare samples will be performed free-of-charge. Please select suitable spare samples able to replace failed samples, enabling your project to continue without delay. We assess the quality of your samples on two parameters, see Table 1.

Table 1. QC parameters for sample assement.

QC parameter	Assessment Method	QC specification
Quantity	Fragment Analyzer / Qubit / PicoGreen / RiboGreen	Depending on your application
Quality	Fragment Analyzer / Agilent BioAnalyzer	RQN ≥ 6 / RIN score ≥7 DQN ≥ 2 (20% high molecular DNA present) Pass Visual Inspection

Quantity

After receipt of your DNA or RNA samples, GenomeScan will accurately measure your DNA and/or RNA sample concentrations by fluorescence based method (Fragment Analyzer / PicoGreen / RiboGreen / Qubit). If the concentrations deviate from the required concentration, we will contact you to discuss further steps.

GenomeScan prefers generally samples with a concentration of 10-100 ng/µl and at least 15 µl. For some applications (see Table 2) more material is required.

Quality

RNA/DNA contamination and degradation are commonly observed and can lead to unsuccessful sequencing experiments. Therefore, it is of upmost importance to assess the sample quality very thoroughly.

The intergrity of DNA samples is assessed on the Advanced Analytical Fragment Analyzer or on agarose gel. The integrity of RNA samples is assessed on the Advanced Analytical Fragment Analyzer or on the Agilent 2100 Bioanalyzer. If the results deviate from the expected (fragment) size, we will contact you to discuss further steps.

DNA/RNA degradation can be detected by a shift in the size distribution towards smaller fragments and a decrease in fluorescence signal of ribosomal peaks (RNA)/high molecular band (10-20kb)(DNA) (see example for RNA degradation in Fig. 4). GenomeScan analyzes your samples with either the Advanced Analytical Fragment Analyzer or the Agilent BioAnalyzer or a combination of the two machines.

In general, the integrity classification ranges from 10 (intact) to 1 (highly degraded). The Fragment Analyzer uses a similar integrity score, the so-called RNA Quality Number (RQN), as given by the Bioanalyzer. The Fragment Analyzer is more sensitive than the Bioanalyzer, since it's RQN score is generally lower than the RIN score calculated by the Bioanalyzer. Your project can only start with samples that have a **RQN score of 6 or higher**. In addition to this RQN score we will **visually inspect** your samples and this serves as a final criterion to pass or fail samples for sample preparation.

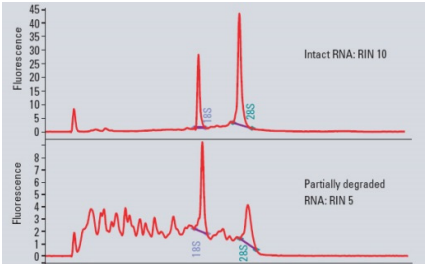


Fig. 4. RNA integrity measurement results after analysis on the Agilent Bioanalyzer: upper picture shows intact RNA with a RIN score of 10; in the middle picture a partially degraded RNA sample is shown which will not pass GenomeScan sample QC; the lower picture shows a strongly degraded RNA sample.

Visual Inspection

GenomeScan performs a visual inspection on all samples. Occasionally, GenomeScan fails a sample that has passed the quality criterion. This is when:

- The whole sample set show mediate RNA quality (RQN scores that fall between ~5.5 and ~6.5). All samples will be flagged as "failed".
- When there is a significant difference in sample quality (RQN 6 vs 10) within a project.
- When there is a significant amount of DNA contamination.
- When the 18S:28S RNA peak is not consistent with other samples in your set.

A sample QC report will be sent to you containing the results, a conclusion and the planning of the next phase of the workflow.

If the sample QC criteria are met (see Table 1), GenomeScan will continue with the experiments without delay. When samples fail one of the two QC criteria (Table 1), GenomeScan can no longer guarantee passing the sample preparation and data QC parameters. You can choose one of the following options:

- Use spare samples to replace the failed samples. Information on the new experimental design is required and will be discussed.
- Use the failed samples and confirm that you agree to process the indicated failed samples in the "QC Entry File" at your Own Risk. The samples are marked as "Own Risk" until the final step of the project.
- Send replacement samples which will delay your project. When choosing this option, you confirm to pay for additional costs that are charged for performing extra sample QCs (€50/sample).
- Leave out the failed samples and continue the project with the remaining samples. This option may have consequences since project materials are already purchased or an extra batch must be initiated.

4.2 Quality Measurement of the Sequencing Run

The main QC method to assess the quality of a flow cell run is the alignment of the PhiX control reads to its reference sequence. This allows quantitative assessment of the error rate for a particular flow cell or lane. Quality scores for individual bases generated in the sample lanes are calibrated against Illumina established standards, giving statistical information regarding the accuracy of base-calling. GenomeScan assesses the overall quality of the run using a defined set of parameters advised by Illumina. The Quality Control report, which is delivered together with the raw data of the project, gives a full description and summary of these parameters.

Chapter 5 Sample Requirements

For each type of sample preparation, specific (optimal) requirements are defined such as type of material, amount of sample. Please read this chapter carefully to ensure that you provide us with the optimal starting material so we can deliver the highest quality output.

We frequently observe that delays in data delivery are caused by diminished sample quality and/or concentration issues.

In general we require the highest quality and purity possible. This can be achieved by multiple methods, however general rules apply. Whenever possible use column or bead-based methods to purify your samples and use non-DEPC treated nuclease-free water as eluant. Elution buffers containing e.g. TRIS/EDTA are not recommended as they can interfere with (spectrophotometric) QCs and downstream procedures. Please ensure that your purification methods are able to adequately remove contaminating RNA from DNA samples and DNA from RNA samples. Please do not employ RNA and or DNA removal techniques after the final elution of your samples without re-purification, as these methods often retain contaminants which heavily interfere with our downstream procedures. This ensures that we can proceed with your experiment without delay.

Table 2. Sample requirements for Illumina NGS. These sample requirements are based on robust input amounts per reaction and the ability to perform QCs and reruns in case of failures. Please note that highly diluted samples are difficult to QC and very concentrated samples lead to relative large losses of material during the QC.

Application	Input material	Optimal amount (per sample)*	Optimal concentration boundaries*
Whole-Genome Sequencing of DNA	Purified and intact gDNA	250-500 ng	5-50 ng/µl
Targeted resequencing of DNA	Purified (PCR) DNA	250-500 ng	5-50 ng/µl
Agilent SureSelect of DNA	Purified gDNA	500/6500 ng	50-200 ng/µl
ChIP-sequencing/complexity reduction of DNA	Enriched DNA	50 ng	>2 ng/µl
Whole Genome Bisulfite Sequencing (WGBS)	Purified gDNA	100 ng	>10 ng
Reduced Representation Bisulfite Sequencing (RRBS)	Purified gDNA	100 ng	>10 ng
mtDNA/Microbiome Sequencing	Purified and intact gDNA	250-500 ng	5-50 ng/µl
16Sv4 Sequencing	Purified and intact gDNA	10-100 ng	>2 ng/µl
RNAseq (Poly-A selection)	Purified total RNA	250-500 ng	5-50 ng/µl
RNAseq (rRNA reduction)	Purified total RNA	250-500 ng	5-50 ng/µl
Small RNA Analysis	Purified, intact total RNA or small RNA enriched equivalent	2500 ng	200-300 ng/µl

**When only lower amounts and/or different concentration series are available for your project contact your Project Manager or Sales Representative to discuss alternative options and/or workflows.*

DNA Analysis services

5.1 (Re)sequencing of DNA – Genomic DNA

Library preparation for (re)sequencing of DNA requires at least 500 ng of purified and intact genomic DNA. To isolate DNA you can choose your own isolation method or any commercially available column-based DNA isolation kit suitable for your tissue type, providing that you mention the isolation kit used on the Sample Submission Form.

5.2 (Re)sequencing of DNA – PCR Products

Instead of genomic DNA, you can also deliver purified PCR products as starting material. Supply us with at least 500 ng purified PCR product and indicate the product size(s). When required, the PCR products will be sheared into smaller fragments for library preparation and sequencing.

RNA Analysis Services

5.3 Whole Transcriptome analysis – mRNA-Seq

Library preparation for RNA-Seq requires purified total RNA. Supply us with at least 500 ng of intact and purified total RNA for Poly-A selection or 500 ng for rRNA-reduction experiments. It is very important to use high-quality RNA as the starting material. Use of degraded RNA can result in lower yield, overrepresentation of the 5' ends of the RNA molecules, or low quality sequencing data.
For RNA isolation you can select any commercially RNA isolation kit available, which is most suitable for your experiment. This is dependent on the source of your RNA, such as type of organism, tissue and tissue quantity. A column purification method which includes a DNAse treatment step is compulsory. In case the sample still contains DNA you need to perform an extra purification step or you can decide to proceed at your own risk. In this case we cannot guarantee our standard data quality.

5.4 Small RNA Discovery

To assess the expression level of small RNAs, supply us with at least 2500 ng of intact and purified total RNA containing small RNAs. Only a small number of RNA isolation kits are able to capture total RNA combined with small RNAs . For example, the mirVana™ miRNA Isolation Kit provided by Ambion and the miRNeasy™ kit provided by Qiagen are kits that enrich fragments smaller than 200 nucleotides. You can also provide us with an isolated fraction of small RNA so we can process this fraction directly.

5.5 FFPE Expression Analysis

Besides the standard mRNA-Seq application, GenomeScan also offers a highly robust method to degraded RNA sequencing solution and therefore it is ideal for profiling and screening formalin-fixed paraffin-embedded (FFPE) samples.

Reduced-representation Analysis Service

5.6 ChIP-Sequencing/Complexity Reduction

The first step in determining which DNA sequences are enriched for DNA:protein complexes is immunoprecipitating the binding site of proteins of interest to DNA. This assay is performed by the customer, after which GenomeScan continues with the sample preparation of the selected DNA fragments.
ChIP samples should be assayed using PicoGreen, which is a ultra-sensitive fluorescent nucleic acid stain for quantitating double-stranded DNA (dsDNA). Nanodrop measurements are not accurate enough at low concentrations and small amounts of contamination may lead to an overestimation of the actual concentration. Please note that there is no single ChIP protocol that will work for every assay. ChIP protocols require optimization according to genome size, genome complexity and the

expected frequency of the binding events being analysed. In literature, many ChIP protocols, optimizing tips and tricks and controls have been described for a wide range of assays, which may be useful for your specific application. Similarly, other (custom) techniques for DNA complexity reduction can also be applied.
GenomeScan has validated low input DNA sequencing down to 2.5 ng, but we recommend to start with at least 50 ng to improve general handling and sequencing quality.

5.7 Methylation Sequencing

Library preparation for WGBS and RRBS requires at least 100 ng of purified and intact genomic DNA. To isolate DNA you can choose your own isolation method or any commercially available column-based DNA isolation kit suitable for your tissue type, providing that you mention the isolation kit used on the Sample Submission Form.

5.8 New Services and PoP Projects

Applications of the Illumina NGS platform are rapidly expanding. Publications on innovative new applications are currently appearing in high impact factor journals on a regular basis. Please contact us for the latest update of our expanding NGS track record or to explore other opportunities.
At GenomeScan, **innovation** is one of our core values; we introduce new technologies, services and applications on a regular basis. Our new services are thoroughly tested before their commercial launch. In advance of each service introduction, we analyze robustness of any new assay, determine quality parameters and deliverables. Furthermore, we make sure our service staff is well-trained, and has the experience and know-how to provide excellent support.
Prior to the introduction of a service, we will open the service to a select number of researchers through our PoP program. PoP experiments are performed in close consultation. For sequencing projects there are possibilities to collaborate on data analysis. Either our in-house staff can analyze your project or you will be invited to analyze the data together with our dedicated Support and Development Specialist at our facility. The data generated with the PoP program will be used to define the quality parameters and deliverables of the new service.

Chapter 6 Delivery of Sequencing Results

After completion of the experimental procedures GenomeScan performs primary data analysis on the sequencing data to ensure that the quality of the experiment is within specifications. The resulting QC and project report is delivered to the customer together with the complete raw output of the sequencer. The most important output file is the sequence file, in which the sequences and their corresponding quality scores are encoded in the common FASTQ sequence format. The Paired-End sequencing data file consists of a second file with the reverse reads that is identical in format and order to the file, in which the sequence data is stored of the reverse reads of the fragments.

6.1 Deliverables of the GenomeScan NGS Service

GenomeScan in general guarantees deliverables for your sequencing project which can be found in Table 3. However, differences may apply depending on the number of samples of your library, the complexity of the starting material, insert size of the library and sequencing methodology like Single-Read or Paired-End sequencing including the required number of bases per read. For Ready-to-Run projects we cannot guarantee a minimal yield per project as we do not control the sample preparation method(s). Please contact your Sales Representative regarding deliverables of your project.

Table 3. Typical deliverables of the Illumina NGS services provided by GenomeScan. Data is given for minimal and maximal number of bases per read. Other options (*e.g.* Paired-End 75) can be discussed with your Sales Representative and/or Project Manager.

Single-Read sequencing	**HiSeq 2500 High Output**	**HiSeq 4000 High Output**	**HiSeq 2500 Rapid Run**		**NextSeq 500 High Output**
Bases per read	50	50	50		75
Clusters passing filter	240*10^6	270*10^6	240*10^6		320*10^6
Reads per lane	240*10^6	270*10^6	240*10^6		320*10^6
Q30 of bases*	≥ 85%	≥ 85%	≥ 85%		≥ 80%
Data output	**12 Gb/Lane**	**13.5Gb/Lane**	**12 Gb/FC**		**24 Gb/FC**
Paired-End sequencing	**HiSeq 2500 High Output**	**HiSeq 4000 High Output**	**HiSeq 2500 Rapid Run**	**NextSeq 500 Mid Output**	**NextSeq 500 High Output**
Bases per read	125	150	250	150	150
Clusters passing filter	240*10^6	270*10^6	240*10^6	100*10^6	320*10^6
Reads per lane	480*10^6	540*10^6	480*10^6	200*10^6	640*10^6
Q30 of bases*	≥ 80%	≥ 75%	≥ 75%	≥ 75%	≥ 75%
Data output	**60 Gb/Lane**	**81 Gb/Lane**	**120 Gb/FC**	**30 Gb/FC**	**96 Gb/FC**

**The percentage Q30 of bases is averaged over the entire run*

Due to unavoidable variation which can result in minor underrepresentation in data yield, a 95% success rate (for project sizes ≥ 20 samples) may apply. All generated data of all samples will be delivered.
The data generated is stored on an external hard disk and will be shipped to you by courier after completion of the project. The data set consists of the FASTQ files extracted by the Illumina software pipeline along with the quality score for each base. The FASTQ files are provided in Sanger FASTQ format and are the starting point for secondary data analysis.
Detailed quality and summary information for each flow cell and each individual sample is provided in our quality assurance report. The quality metrics are based on the specifications provided by Illumina and GenomeScan specification for a good run. For targeted sequencing using the Agilent SureSelect Target Enrichment the percentage 30X coverage is also provided per sample.

The data is also accompanied by a project report which reports the workflow and procedures that were followed, the results of the (intermediate) QCs that have been performed, and the results of the sequencing run(s). A document that describes the file formats and interpretation of the data is provided as well.

GenomeScan does not include raw image files and intermediate files generated during processing of raw signal intensities and base-calling due to the size of the data. These data are removed from our servers immediately after the Illumina software pipeline has processed the data successfully.

6.2 Data Analysis Options

The amount of data generated with NGS experiments is large. Handling of this data requires extensive capacity of computing resources as well as bioinformatics skills. Therefore, you can visit GenomeScan after completion of your project for a meeting during which the results of your project results are discussed. Furthermore, this meeting allows us to directly answer questions and show you how you can proceed with your data analysis or the possibilities for data analysis by our bioinformatics department.
Of course, we can perform extended data-analysis for you through our bioinformatics program. Please contact us for our Next Generation Data Analysis Guideline which contains predefined analysis packages or for a tailor-made solution to answer your specific research question(s).

Chapter 7 Sample Policy and Requirements

In the following paragraphs a brief outline is given of what you can expect from GenomeScan regarding our policies.

7.1 Sample Shipment

Before sending your samples to GenomeScan, please notify your Project Manager so he/she can advise you when to send your samples to ensure that the shipment is on time for the scheduled start of the project. We can also help you avoid delays such as during (Dutch) bank holidays. Do not ship samples just before the weekend, but preferably at the start of the workweek. Make sure your Project Manager knows the delivery date, so we can track shipment if necessary.

Ship all samples in 96-well plates labelled with your project code or in vials labelled with our unique GenomeScan codes (GS_ID) as indicated on the Sample Submission Form. When plates are used also confirm the corresponding plate locations (A1-H12) with your samples as indicates on the Sample Submission Form. Send your samples to us in a sealed bag or box in a polystyrene container. GenomeScan recommends that (g)DNA are sent on ice-packs to ensure that the sample quality remains stable. RNA samples should be sent on dry ice. Take into account that shipment may take longer than expected, especially in case of international shipping, and that the package contains sufficient dry ice.

7.2 Sample Delivery

Upon receipt of the signed PO form, we ask our customers to give us an estimation of the delivery date of the samples. If this is not known at that time, please notify us as soon as possible by sending us the completed Sample Submission Form with the correct date. We use this date to schedule the first step of the experimental workflow (see chapter 2). When the delivery of your samples is delayed, inform us as soon as possible so we can reschedule your project. For rescheduling of your project costs may be charged for lost time slots.

7.3 Ordering of Project Specific Perishable Reagents

GenomeScan will order materials and reagents upon receipt of the signed PO form when not available, just before the sample delivery date that you have indicated on the Sample Submission Form. Certain projects require materials or reagents that are solely used during your specific project. If this is the case, you will be informed of the expiration date. It is your responsibility to deliver samples well within the expiry date of these reagents. GenomeScan cannot use reagents or materials past the expiry date as part of our requirements for ISO 17025 certification.

7.4 Batch Policy

The number of batches is agreed upon and documented in the quotation. When deviating from the quotation additional costs may be charged.

7.5 Spare and Replacement Sample Policy

The quality of samples will be analyzed at the start of the project. GenomeScan gives you the option to deliver 10% surplus samples. The QC assessment for the spare samples is free-of-charge. In case one or more samples do not meet our QC requirements, we inform you of this result. It is our standard procedure to replace failed samples by spare samples without delay of the project. When replacement of failed samples with spare samples is not possible, you may send a replacement batch to

GenomeScan. Extra costs apply for replacement samples to cover for the extra QC assessments and rescheduling of your project. Furthermore, sending replacement sample(s) usually affects the turnaround time of your project.

7.6 Use of 'failed' Samples

Samples, especially those consisting of RNA, are prone to fail the sample QC. You can choose to continue with these samples when you agree to process the failed samples at your own risk. GenomeScan will no longer guarantee high data quality standards or any sample preparation results for those samples that are labeled 'Own Risk'. GenomeScan will offer this option for samples for which GenomeScan deems that there is considerable chance that processing will be unsuccessful or result in non-usable data. You will be asked to confirm that you agree to process the failed samples at your own risk. The quality of the validated control sample PhiX that is taken along during each the run, will be used to assess the performance of GenomeScan. If no technical failure is detected, we conclude that the run is within specifications.

7.7 Project Delays

GenomeScan is dependent on its suppliers for materials and technical support. When delays are unavoidable, GenomeScan will notify you as soon as possible and keep you updated on the progress.

7.8 Use of Alternative Devices

GenomeScan reserves the right to perform specific parts of projects such as the NGS run on devices located in the laboratories of our partner, the Leiden Genome Technology Centre. These experiments remain performed under strict ISO/IEC 17025 guidance. This guarantees that no customer or sample information is disclosed.

7.9 Sample Storage Terms

Samples present at GenomeScan are stored for 6 months after completion of your project. On the Sample Submission Form sent to you at the start of the project, you can choose to receive the samples after the project is finished or that the samples will be stored at GenomeScan and discarded after the indicated period. If the samples are returned to you by courier, additional shipment costs apply.

7.10 Data Storage Terms

The sequence files with quality information will be sent to you at the end of your project. Because of storage limitations we cannot maintain a readily available copy on our data servers. Therefore, we strongly advise to check if you have correctly received the data of all samples after receiving the hard disk.

Appendix Abbreviation List and Glossary

ChIP	Chromatin Immunoprecipitation
CNV	Copy Number Variation
PoP	Proof-of-Principle; Type of project used for introducing new protocols/technologies pre-launch
FASTQ	A text-based format containing the base sequence with its corresponding Q-score
FFPE	Formalin-Fixed Paraffin-Embedded
Index	Unique redundantly coded identifier tag, also called adapter sequences
IP	Immunoprecipitation
Multiplex	Sequencing mode which allows the pooling of samples into one lane (or more)
NGS	Next Generation Sequencing
PE	Paired-End
PM	Project Manager
PO	Purchase Order
QC	Quality Control
QMS	Quality Management System
RAD	Restriction site Associated DNA
RIN	RNA Integrity Number: a score to indicate the RNA degradation by the Bioanalyzer
RQN	RNA Quality Number: a score to indicate the RNA degradation by the Fragment Analyzer
SBS	Sequencing-by-Synthesis
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure
SR	Single Read
SSF	Sample Submission Form
WGBS	Whole Genome Bisulfite Sequencing
RRBS	Reduced Representation Bisulfite Sequencing

Caring for your future