



# Sample Preparation and Sequencing for the Illumina Genome Analyzer Ix

Donna-Lee Fancy, Darrin Klassen, Kim  
Nelson, Christine Sidebottom,  
Carling Tallon

January 20, 2011



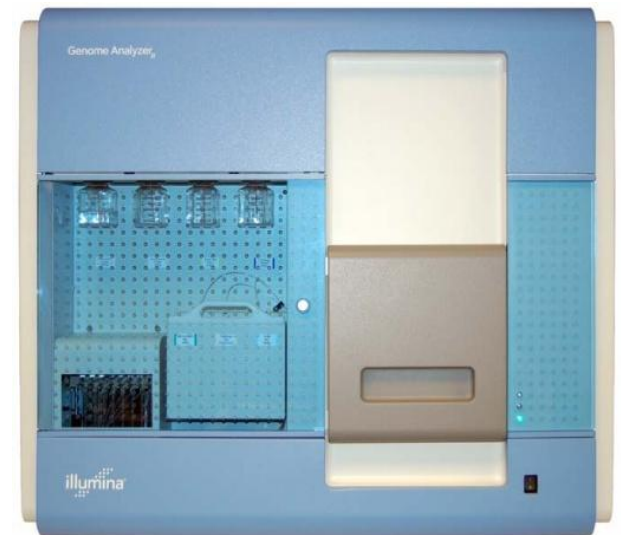
National Research  
Council Canada

Conseil national  
de recherches Canada

Canada

# Illumina GAllx

- 1<sup>st</sup> Illumina Genome Analyzer installed at PBI June 2009
- 2<sup>nd</sup> Illumina GA installed March 2010 (AAFC machine)
- Provide sequencing support for Internal projects, Fee-For-Service, and other projects such as CanSeq, GHI, TUFGEN



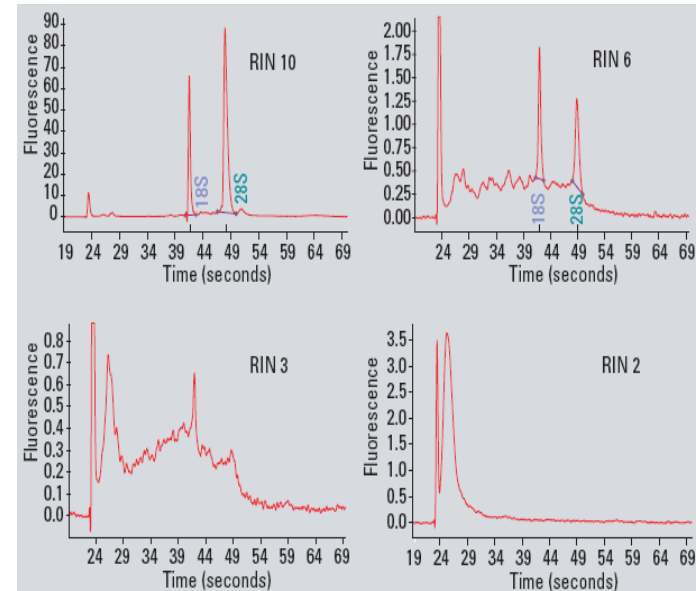
# Current Applications

- Starting material – Genomic DNA, cDNA, Total RNA, mRNA, enriched small RNA, Genomiphi
- Single Read – 36 to 150 bp, one direction only
- Paired End – 36 to 150 bp, both directions, gap size approximately 100 to 200 bp
- Mate Pair – 36 bp, both direction, gap size range from 2 to 5 kb
- Multiplexing up to 12 samples per lane
- miRNA – 36 bp, one direction

# Sample Submission

- A. Contact Sequencing facility to coordinate contract set up with business office and account creation in Laboratory Information Management System (LIMS)
- B. Enter sample information into LIMS
- C. Label your tube as it appears in LIMS, include Request ID #, contract holders name and deliver material to PBI
- D. Sample quality is checked:

- Genomic DNA quantity checked using fluorometric quantification - Qubit
- RNA check with Bioanalyzer for quantity and quality.  
Total RNA should have a RNA Integrity number (RIN) greater than 8
- Please ensure you are submitting high quality material and have quantified the sample
- Contact me for current quantities and volumes required - new kits are currently being released which require lower quantity of input material



Agilent Bioanalyzer RNA electropherogram

# Illumina Sequencing Overview – 3 Steps

## 1. Library Preparation

Prepares DNA and RNA for ligation of adaptors



Prepackaged Library Preparation Kits

## 2. Cluster Amplification

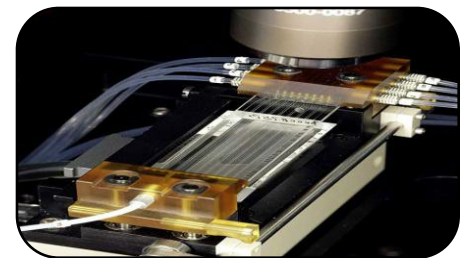
DNA hybridized to flowcell and cluster amplification is complete using cBOT



Illumina cBOT

## 3. Sequencing on the Genome Analyzer

-Imaging of the Clusters



Flowcell in position in Genome Analyzer

# 1. Library Preparation

- A. Production of Fragmented DNA
  - Genomic DNA Nebulized to produce to 200 – 800 bp fragments
  - mRNA isolated from Total RNA using oligo dT beads, chemical shearing, cDNA synthesis
  - Mate Pair – Genomic DNA sheared, ends Biotin labeled, gel size selected 2 to 5 kb, circularize DNA, Fragment, extract Biotin labeled fragments.
- B. Repair ends / Add A overhang
- C. Ligate Adaptors – Single Read or Paired End Read
- D. Gel Size Select Ligated DNA – 200 to 500 bp
- E. PCR Enrichment
- F. Bioanalyzer check for size/quantity – qPCR option

# New for 2010 SPRIworks

**SPRI-TE Nucleic Acid Extractor**

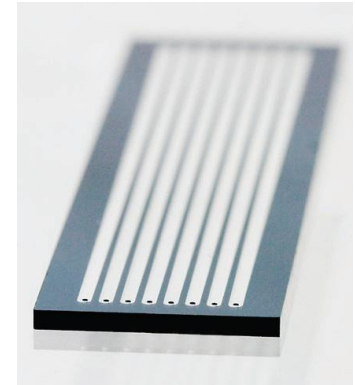


- Fully automated fragment library system for the Illumina and 454 sequencers
- Allows 5x more NGS libraries to be processed over manual library construction – 20 per day
- Solid Phase Reversible Immobilization (SPRI) paramagnetic bead based technology
- Reproducible and consistent library output
- Improved sequencing service turnaround time

# 2. Cluster Generation on flowcell using cBOT

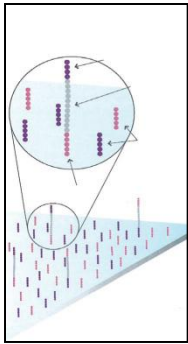
## • Flowcell

- contains 8 Lanes or Channels – One lane reserved for phiX control
- Contains “lawn” of oligos arranged in tiles
- Each lane contains 2 rows of x 60 columns of tiles each 0.5 mm
- Each tile can generate approximately 300k of clusters

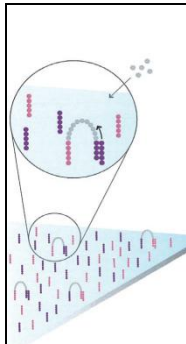


## • cBOT

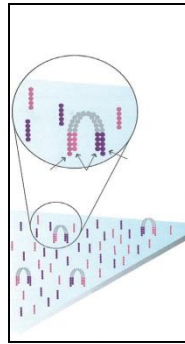
- Automated instrument used for cluster generation
- Reagents supplied in 96 well plate
- Requires approximately 15 minutes hands on time, 4.5 hours to completed flowcell



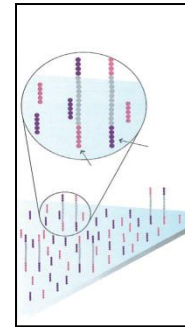
**1.** Hybridize DNA to the surface of the Flowcell



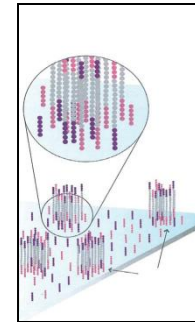
**2.** Bridge Amplification



**3.** Single strand becomes double stranded DNA



**4.** Strands are denatured and bridge amplification continues



**5.** Cluster formation complete – Reverse stands cleaved – Ends Blocked – Seq. primer annealed

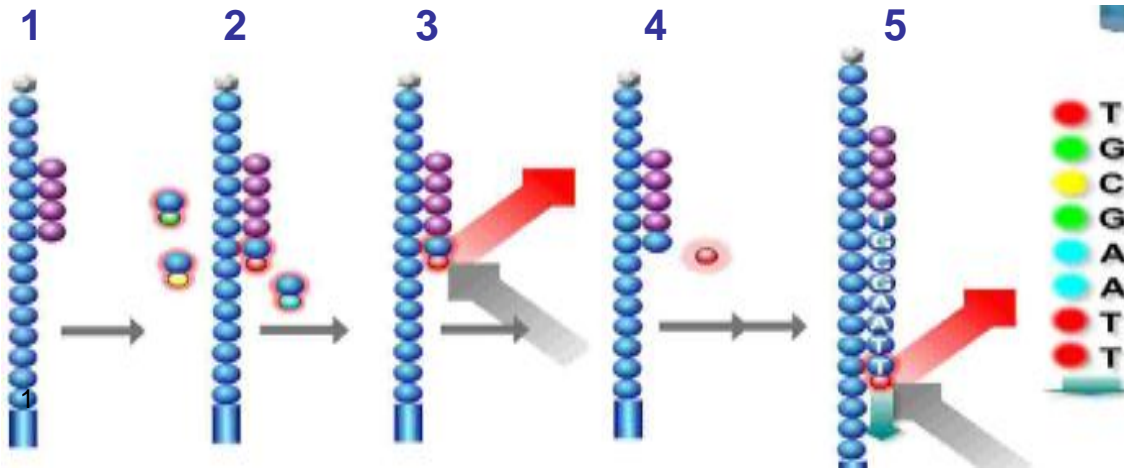
# 3. Sequencing on the GAllx

- The GA images clustered flowcells using Sequencing by Synthesis (SBS)
- Uses reversible terminators with removable fluorescence, which is not as sensitive to homopolymeric regions
- During imaging the GA performs Real-Time data analysis, which reduces the amount of data transferred and speeds up further analysis, lowers data storage
- GA can regenerate the flowcell for Paired-End reads while the flowcell remains in position on GA using the Paired-End Module
- GA requires 70 minutes to incorporate and image flowcell for each base pair addition



Paired end Module

# Imaging on the GA



- 1 - Flowcell loaded into GA
- 2 - First reversible dye terminator incorporated
- 3 - Two lasers cause excitation of the labeled NTP and image is recorded - First base report generated and evaluated for cluster numbers, focus quality, and image intensity
- 4 - Fluorescent terminator is cleaved
- 5 - Incorporation continues for the number of cycles selected

## Genome Analyzer<sub>IIx</sub> Performance Parameters

Read Length	Run Time (Days)	Output (Gb)
1 x 35 bp	~2	10 - 12
2 x 50 bp	~5	25 - 30
2 x 75 bp	~7	37.5 - 45
2 x 100 bp	~9.5	54 - 60
2 x 150 bp	~14	85 - 95

\*Sequencing output generated using TruSeq SBS V5 kit with PhIX library and cluster densities between 508,000-630,000 clusters/mm<sup>2</sup> that pass filtering on a GA<sub>IIx</sub>.

### Throughput

Up to 6.5 Gb per day for a 2 x 100 bp run

### Reads

Up to 320 million clusters passing filter and up to 640 million paired-end reads

### Performance

The Genome Analyzer<sub>IIx</sub> generates a significant yield of bases greater than Q30

- Greater than 90% bases higher than Q30 at 2 x 50 bp
- Greater than 85% bases higher than Q30 at 2 x 100 bp

### Service and Support

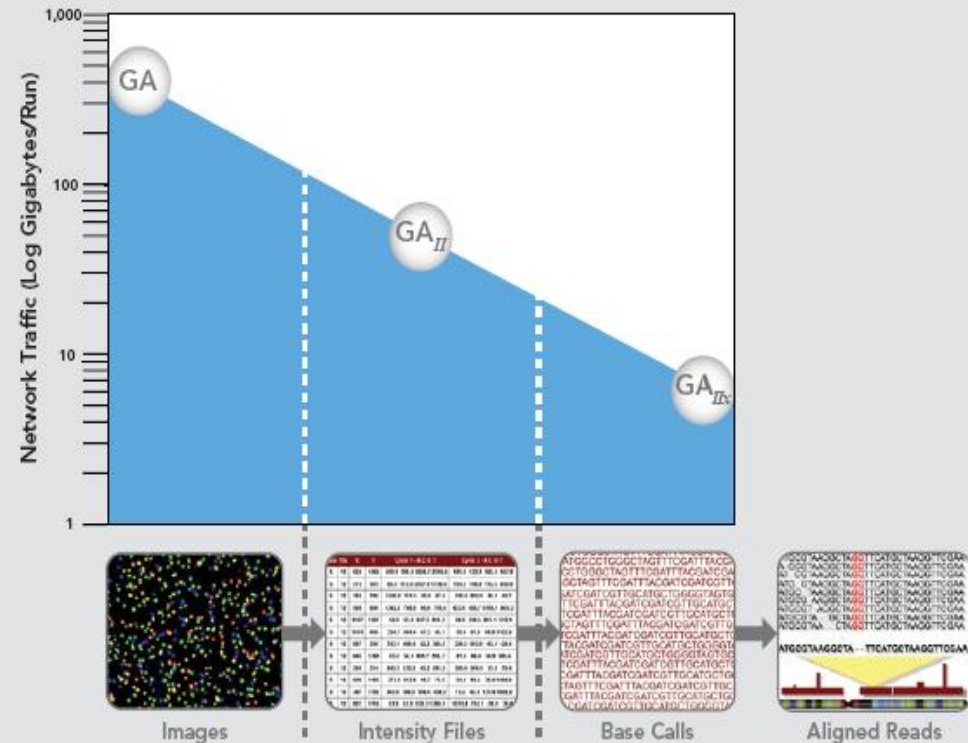
Illumina will ensure that your Genome Analyzer<sub>IIx</sub> is properly installed and qualified, and will provide ongoing maintenance and service. This industry-leading support is available in North America, Europe, and Asia

# Data Transfer and Analysis

## Real-Time Analysis (RTA)

- Converts raw image files into intensity files
- Deletes image files (7 + TB)
- Converts intensity files into basecalls
- Exports intensities and basecalls to Bioinformatics for further analysis

FIGURE 2: DECREASING NETWORK BURDEN WITH SYSTEM ADVANCES



Real-time analysis generates base call-level data instead of bulky images or intensity files. This is performed on the instrument workstation, reducing the need for large file transfer over networks and complex data storage infrastructure.

# Future Upgrades

## Illumina HiSeq 2000

Scheduled to arrive mid-March 2011

- **Dual Flow Cells** – Sequence a single or dual flowcell with different read lengths simultaneously
- **Dual Surface Imaging** – Top and bottom of flowcell are imaged using 4 camera system

**High-Capacity Reagent Chiller** – Holds enough reagent to sequence up to 209 cycles per flowcell

**Integrated fluidics for Paired-End Runs** – Stored in reagent compartment

**Oil Free Imaging** – Flowcell held in place using vacuum

**600Gb throughput by Summer 2011**



# LIMS - Login

Fiesta :: NRC-PBI Lims Home



## NRC-PBI Sequencing LIMS

Logged in as **Al Dente** (*aldente*)

[Settings](#) [Logout](#)

### Introduction

Welcome to the NRC-PBI Sequencing Management System. Use the links on the right to request next generation sequencing services, to track the status of existing requests, and to download the results of completed sequencing runs.

To inquire about our next generation sequencing services please contact [Dr. Andrew Sharpe](#).

For more information about the available sequencing platforms, please follow the links below.

[Roche 454 Genome Sequencer FLX Titanium](#)

[Illumina Genome Analyzer Iix](#)

#### **Sequencing Request Form**

Fill out this form to request a new sequencing job. You must have a valid sequencing contract number to complete this form.

#### **Browse by Sequencing Requests**

View the status of past and pending sequencing requests.

#### **Browse by Sequencing Runs**

View sequencing runs that include your samples.

# LIMS – Sequencing Request Form

Fiesta +

NRC-PBI LIMS Sequencing Request Form

Welcome to the NRC-PBI Sequencing Management System. The form below can be used to request sequencing services.

Please note that service agreements are handled independently and must be arranged prior to sequencing. To arrange a fee-for-service contract please contact [Dr. Andrew Sharpe](#). For internal projects please contact your Project Manager to identify the appropriate program name. Please fill out the form as completely as possible. All fields are required.

## Sequencing Request Form

**Principal Investigator:**

**Sequencing Platform:**

**Funding Type:**  internal  fee-for-service

**Program:**

Select "internal" if the sequencing is part of an existing PBI project, then select the project from the drop-down list.

Select "fee-for-service" otherwise, and then enter the contract number.

If you have any questions about funding please contact Dr. Andrew Sharpe.

# LIMS – Sequencing Request Form

## NRC-PBI LIMS Sequencing Request Form

### Sequencing Request Form

**Run Protocol:** mRNA

**Multiplexed?:**  No  Yes

**Number of Materials:** 2

**Number of Cycles:** 100

**Direction:** Paired End

**Material Type:** Total RNA

#### Material 1 of 2

**Material Name:** BXR-1

**Number of Lanes:** 2

**Organism:** Arabidopsis thaliana

**Genotype:** **Arabidopsis thaliana**  
mouse-ear cress

**Tissue:** NCBI TaxonId: 3702

**Treatment:** n/a

**Quantity of DNA (ng/μL):** 100

**Comments:**

Enter the scientific name of the organism to be sequenced. The name must match an entry in [NCBI's Taxonomy database](#).

#### Notes

Some organisms have more than one scientific name in use. If the name you are using is not accepted, please go to the [NCBI Taxonomy database](#) and attempt to find the name that NCBI considers to be the "correct" name.

Many types of metagenomic samples are also included in the NCBI Taxonomy database, and may be entered here. For example "soil metagenome" or "human gut metagenome". For a list of available metagenomes please see [NCBI Taxonomy's Unclassified sequences page](#)

If you are certain that your organism is not available in NCBI's Taxonomy database under any name, type "unidentified" in this field and provide a detailed description of the organism/metagenome in the comments field below.

# LIMS – Browse by Sequencing Request

Fiesta :: Login    Fiesta :: Requests    Open a new tab    Requests    Logged in as Al Dente (aldente)    Settings    Logout

Requests

Search field Request Name for    Submit Query     Show already sequenced requests

<< Previous Page    Showing results 1 - 2 of 2    Next Page >>

Request Name	Requested By	Principal Investigator	Contract Number	Program	Request Time	Platform	Number of Materials	Status
R000DB	Al Dente	Dr. Al Dente	2010-5555	External	2010-09-24 15:16:24	Illumina GAII	1	processing
R000DA	Al Dente	Dr. al dente	2000-5555	External	2010-09-24 15:10:51	Illumina GAII	1	unassigned

## Request Status Legend

Status	Description
unassigned	All materials for this request are unassigned to any sequencing run
processing	Materials have been assigned to sequencing runs, but all sequencing runs are not yet completed
completed	All samples for all materials have been fully sequenced

# LIMS – Browse By Sequencing Run

Fiesta :: Sequencing Runs

## Sequencing Runs

Logged in as [redacted] [Settings](#) [Logout](#)

Sequencing Runs

Search  for    Show completed runs

<< Previous Page      Showing results 1 - 3 of 3      Next Page >>

Run Name	Plate Name	Platform	Operator	Creation Date	Description	Status
100928_B_00010_FC	61G7YAAXX	Illumina GII	Darrin Klassen	2011-01-17		completed
A-2010-04-12	n/a	Illumina GII	Darrin Klassen	2010-12-03		completed
100806_A_00007	61T3HAAXX	Illumina GII	Darrin Klassen	2010-11-26		completed

## Status Legend

Status	Description
requested	A sequencing run record has been created by the sequencing unit, but actual sequencing has not completed.
processed	Actual sequencing has finished, but bioinformatics post-processing has not
completed	Final stage when sequencing and bioinformatics processing have completed. Note that this includes only processing of raw data, not any downstream analyses such as assembly or annotation.

# LIMS – Browse By Sequencing Run

Fiesta :: Sequencing Run 100928\_B\_... +

Sequencing Run 100928\_B\_00010\_FC Logged in as [redacted] [Settings](#) [Logout](#)

### Run Information

Run Name: 100928\_B\_00010\_FC  
Machine Name: B  
Plate Name: 61G7YAAXX  
Platform: Illumina GII  
Run Type: single-end  
Num Cycles: 100  
Operator: klassend  
Status: **completed**  
Creation Date: Mon Jan 17 15:41:59 2011  
Description:

Lane	Sample Name	Biomaterial	Biomaterial Type	Principal Investigator	MID	NumReads	Download
3	[redacted]	[redacted]	Genomic	[redacted]	n/a	4,098,613	

### Run Browser

Click [here](#) to download the run statistics or select a lane to display: Lane 3

#### Quality Scores for Run\_100928\_B\_00010\_FC\_Lane\_3\_1

Quality Score (40=Highest, 2=Lowest)

Quartiles (red lines)  
Medians (black line)

#### Nucleotides distribution for Run\_100928\_B\_00010\_FC\_Lane\_3\_1

% of total (per read position)

Legend: N (pink), T (yellow), G (green), C (red), A (blue)



Thank you!



National Research  
Council Canada

Conseil national  
de recherches Canada

Canada