EasyGenomics

华大基因 BGI

EasyGenomics

Next Generation Bioinformatics on the Cloud

http://www.easygenomics.com

Sifei He
Director of BGI Cloud
hesifei@genomics.cn

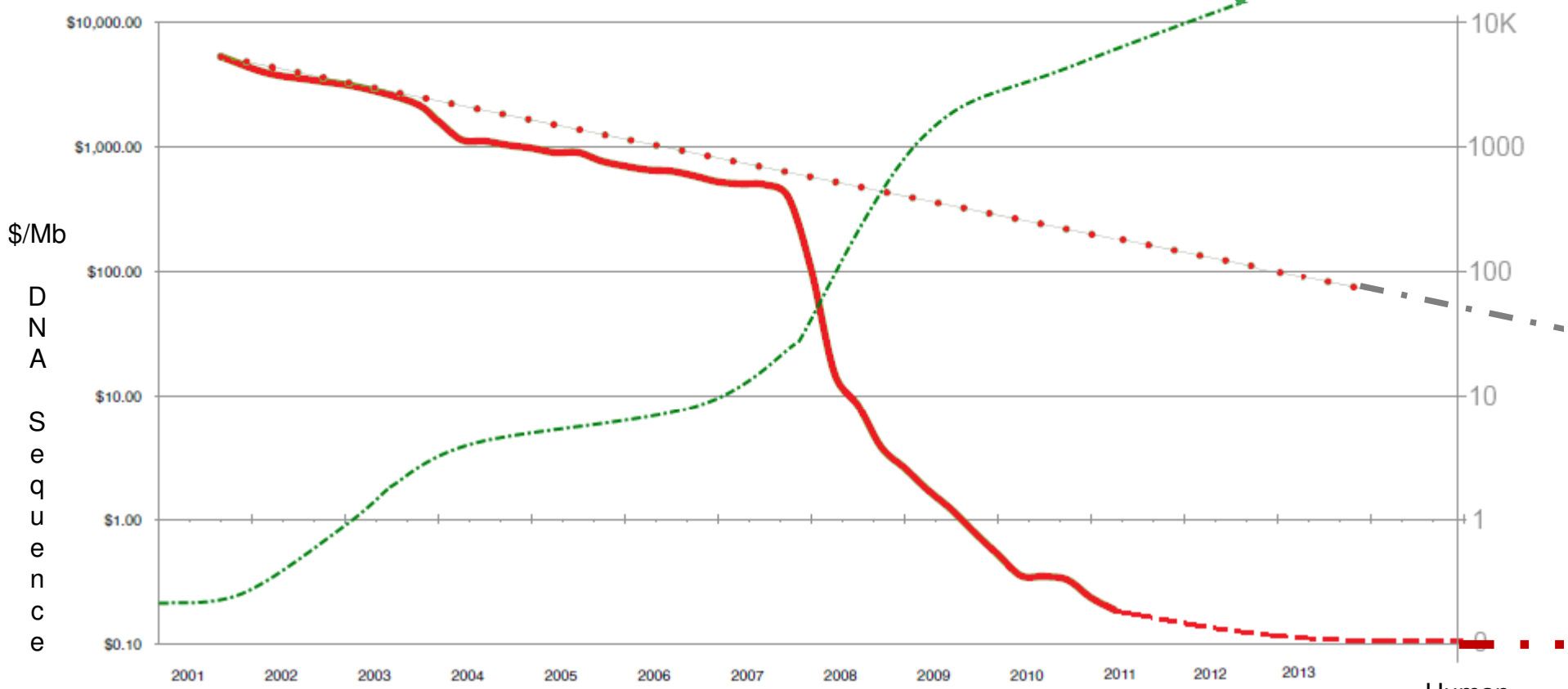
Xing Xu, Ph.D
Senior Product Manager
EasyGenomics | BGI
xuxing@genomics.cn

Contact Us
info@easygenomics.com

Agenda

- Vision and Strategy
- Problems and Solutions
- Product Introduction
- LIVE Demo
- Future Roadmap
- Q&A

Trend of Volume and Cost

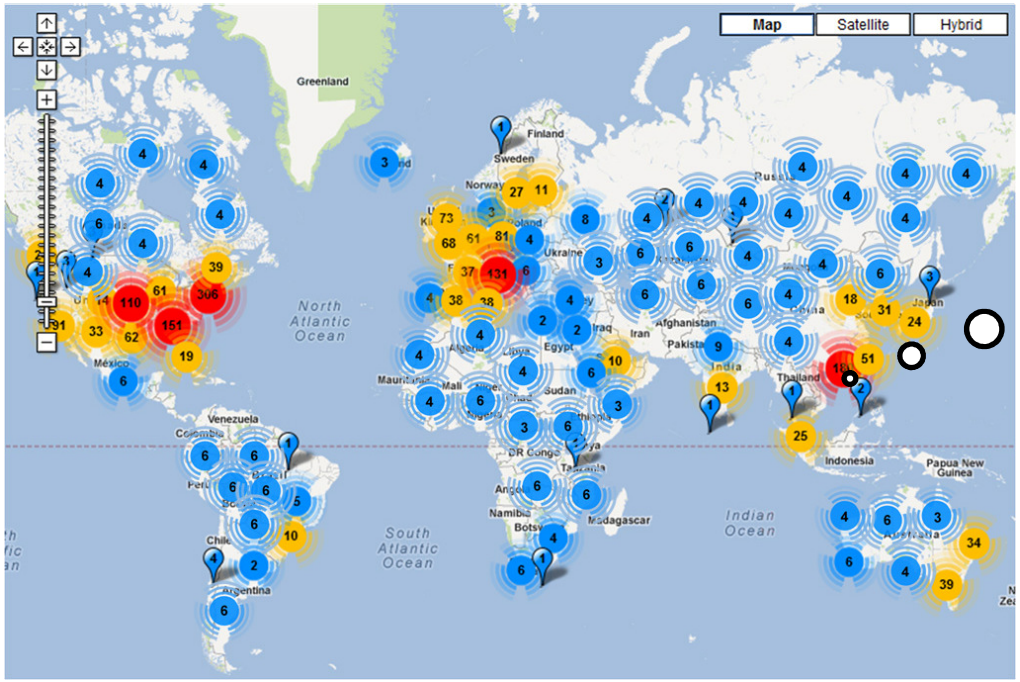


Figures adapted from Sboner A, *et al.*: The real cost of sequencing: higher than you think! *Genome Biology* 2011, 12:125
Numbers and Images from private research and the open Internet

Geological side of the problem

+

Sequencing is a COMMODITY and happens EVERYWHERE.

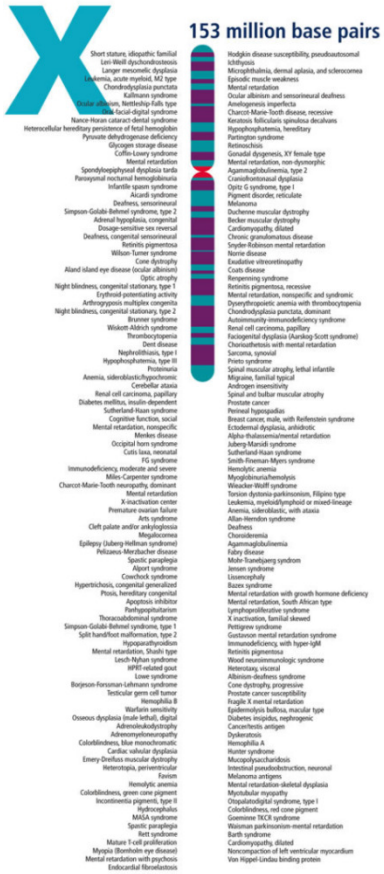


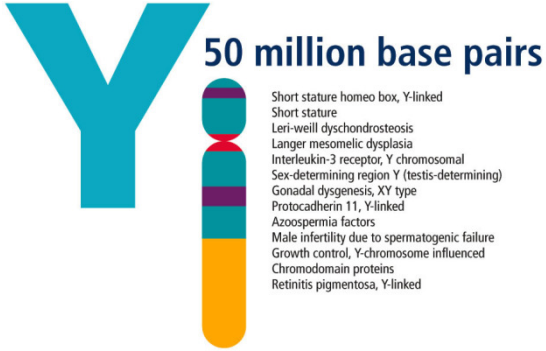
BGI

Images from omicsmaps.com

Interpretation is the KEY

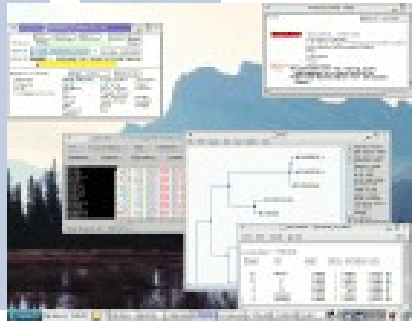
- **Analysis** and **Interpretation** is the KEY
- **Application** is the "Silver Bullet"





Difficulties of Analysis

Primary analysis	Secondary Analysis	Tertiary Analysis	Post Tertiary Analysis
Base calling	Mapping	Variant Calling	In-depth Annotation
Data throughput	Computation intensive	Complicated Algorithms	Lack of knowledge
Data storage	Data storage	Computation intensive	



Problems and Solutions

Solutions

Cloud

High Speed Data Exchange

Workflows

+) Resource Management

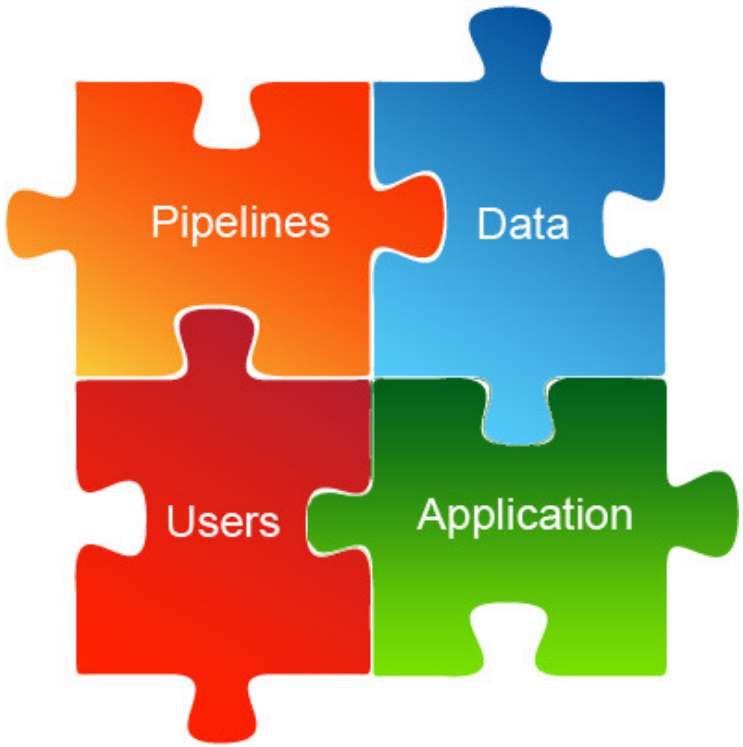
EasyGenomics

Problems:

- ~~Big genomic data~~
- ~~Geological distribution~~
- ~~Algorithm integration~~
- ~~Computational demand~~

EasyGenomics™

- **EasyGenomics is the bioinformatics platform for research and applications on the cloud**



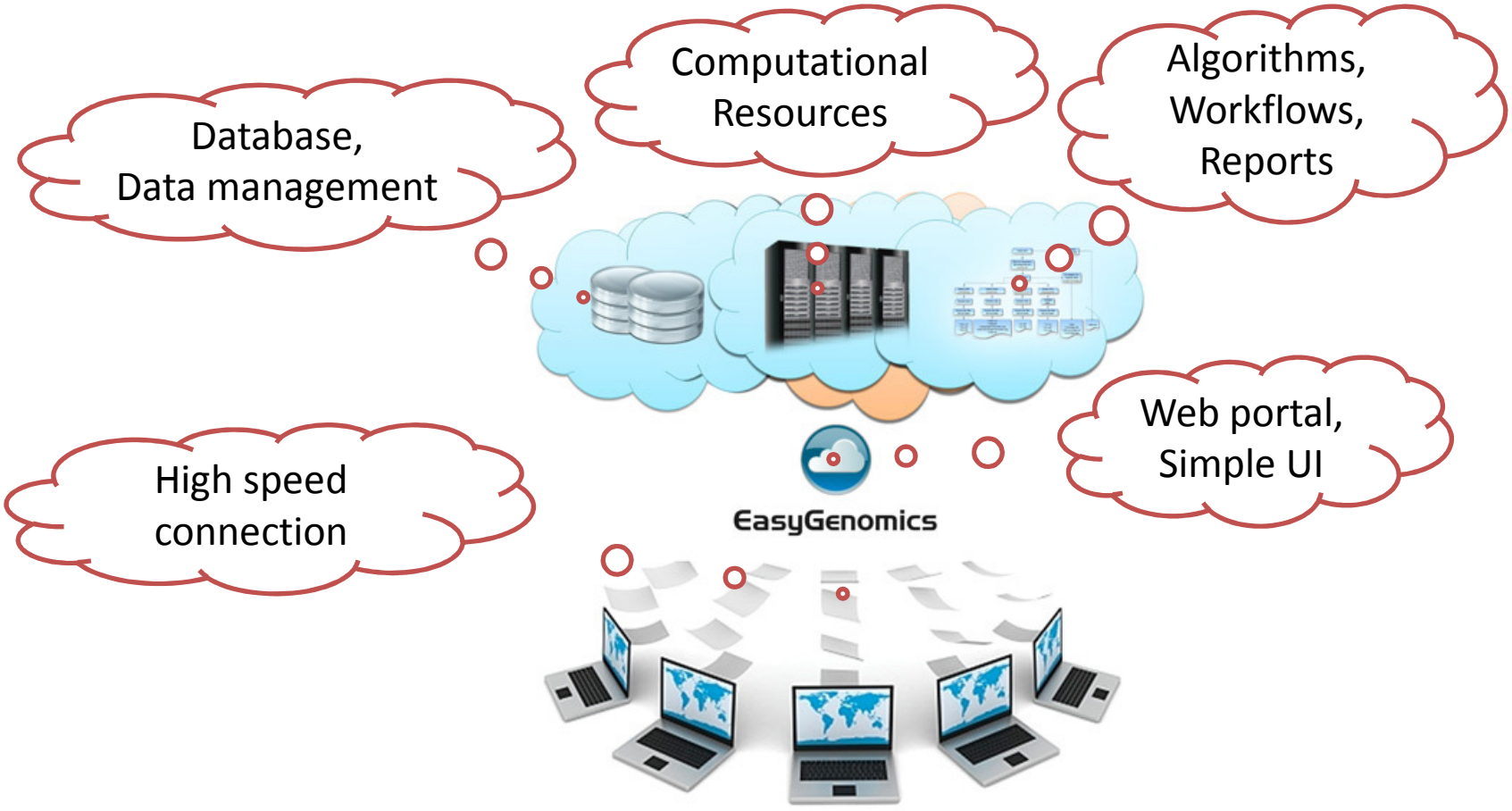
High speed data exchange

Fast and accurate pipelines

Easy and straightforward UI

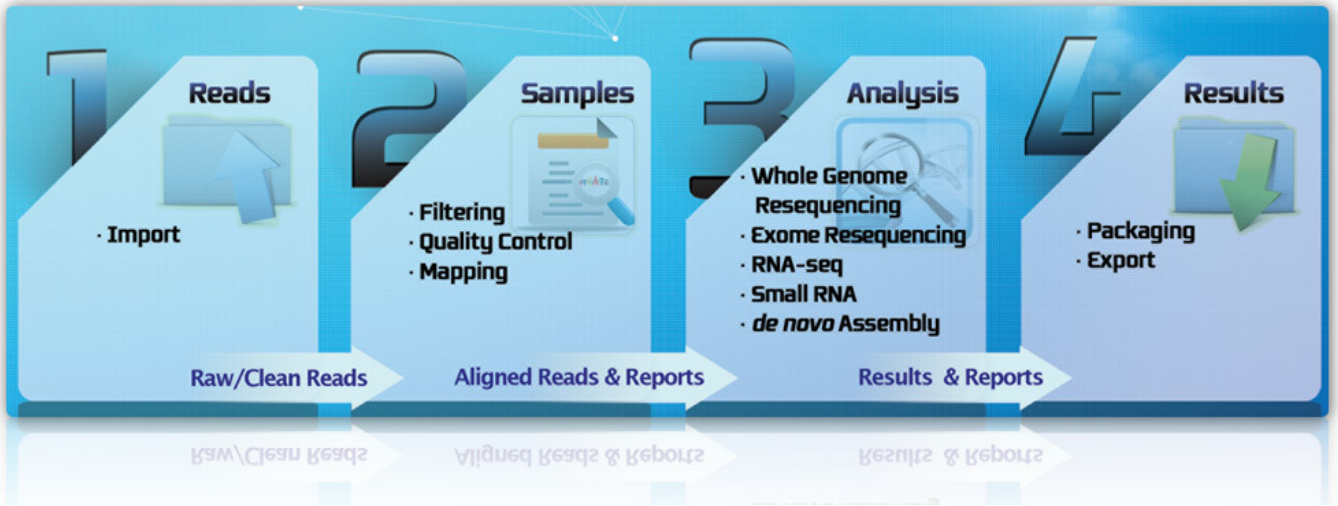
Application integration

EasyGenomics™



EasyGenomics is the bioinformatics platform for research and applications on the cloud

Bioinformatics Core



- **Algorithms:**

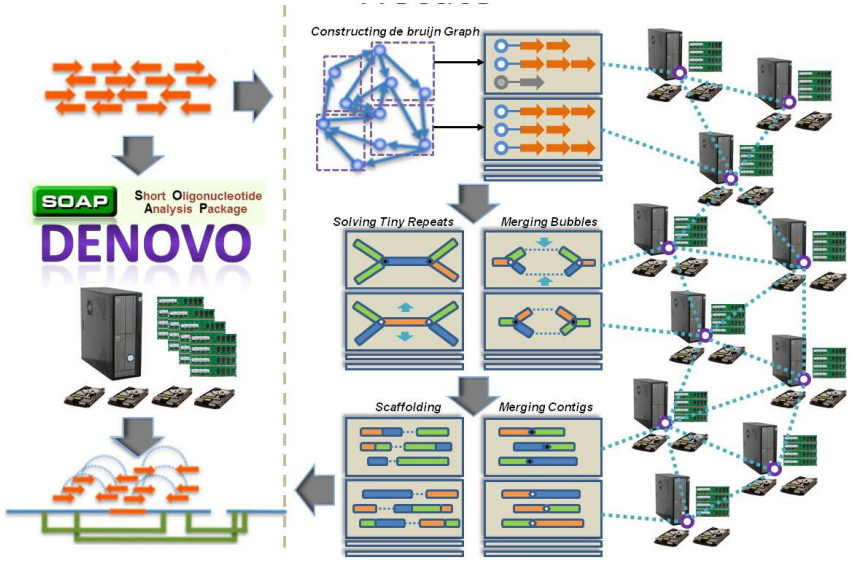
  Carefully chosen, tested and optimized

- **Workflows:**

  Whole genome resequencing, exome resequencing, RNA-Seq, small RNA, *de novo* Assembly

Enabling Technology

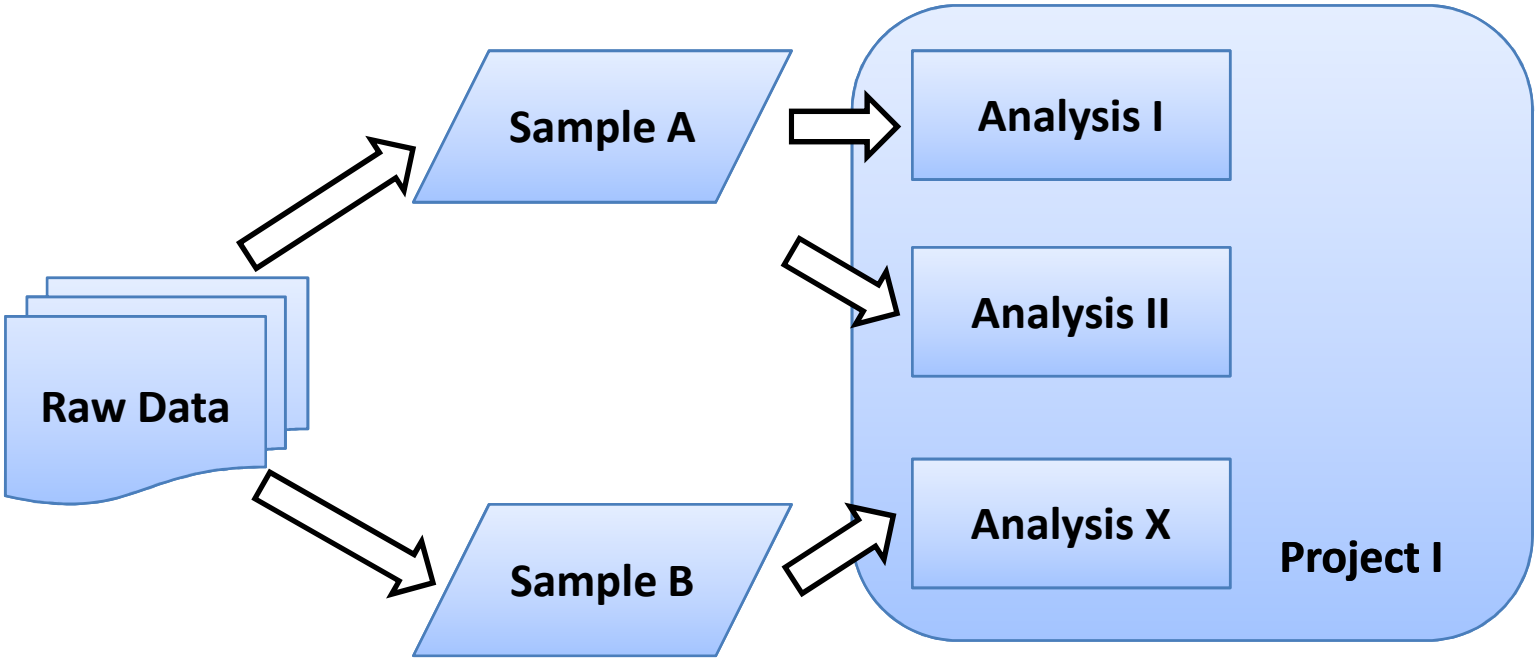
Best Practice Award for IT Infrastructure



Hadoop-based Flexible Computing

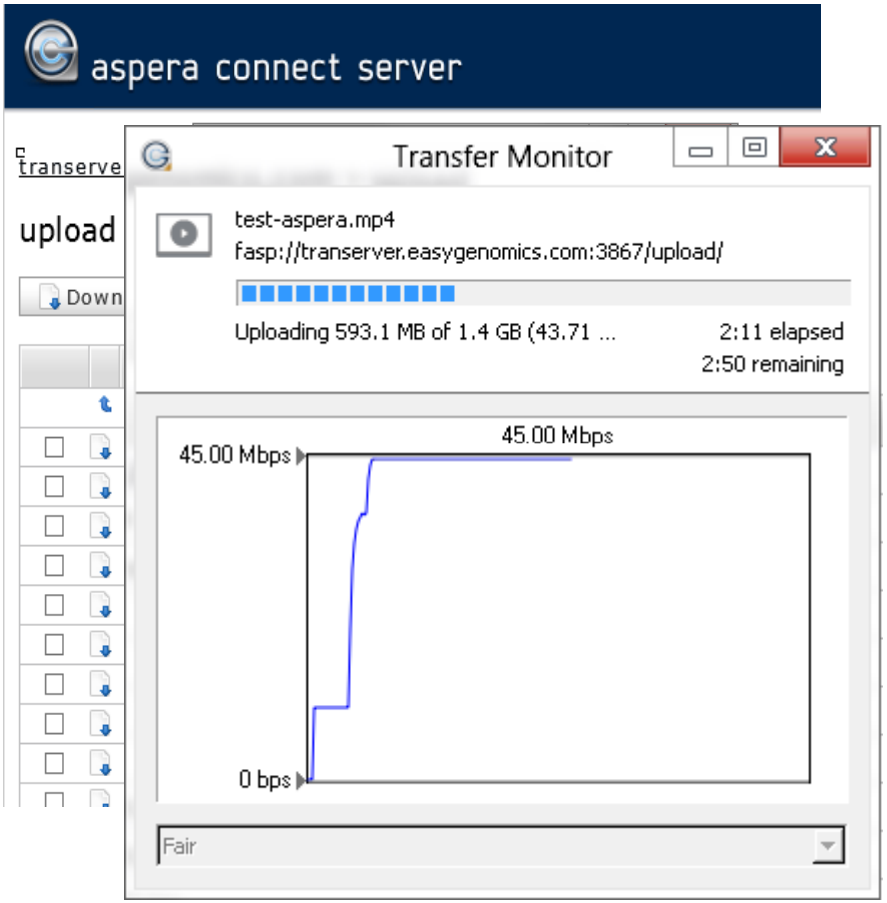
Human Genome	SOAP*denovo*	EasyGenomics™ (192 cores)
Genome Coverage	86%	86%
Assembly Time	70h	55h
No. of Servers	1	15
Memory Size	500GB x 1	24 GB x 15
Mode	Centralized	Distributed

Data Management



- **"Sample", "Analysis", "Project"**
- **Mimicking real research procedure**
- **Automatic management of underlying data structure**

High Speed Data Exchange



- **Aspera's patented *fasp™* high-speed file transferring technology**
- **10~100X faster than FTP**

Resource Management

Multitenancy

Workspace

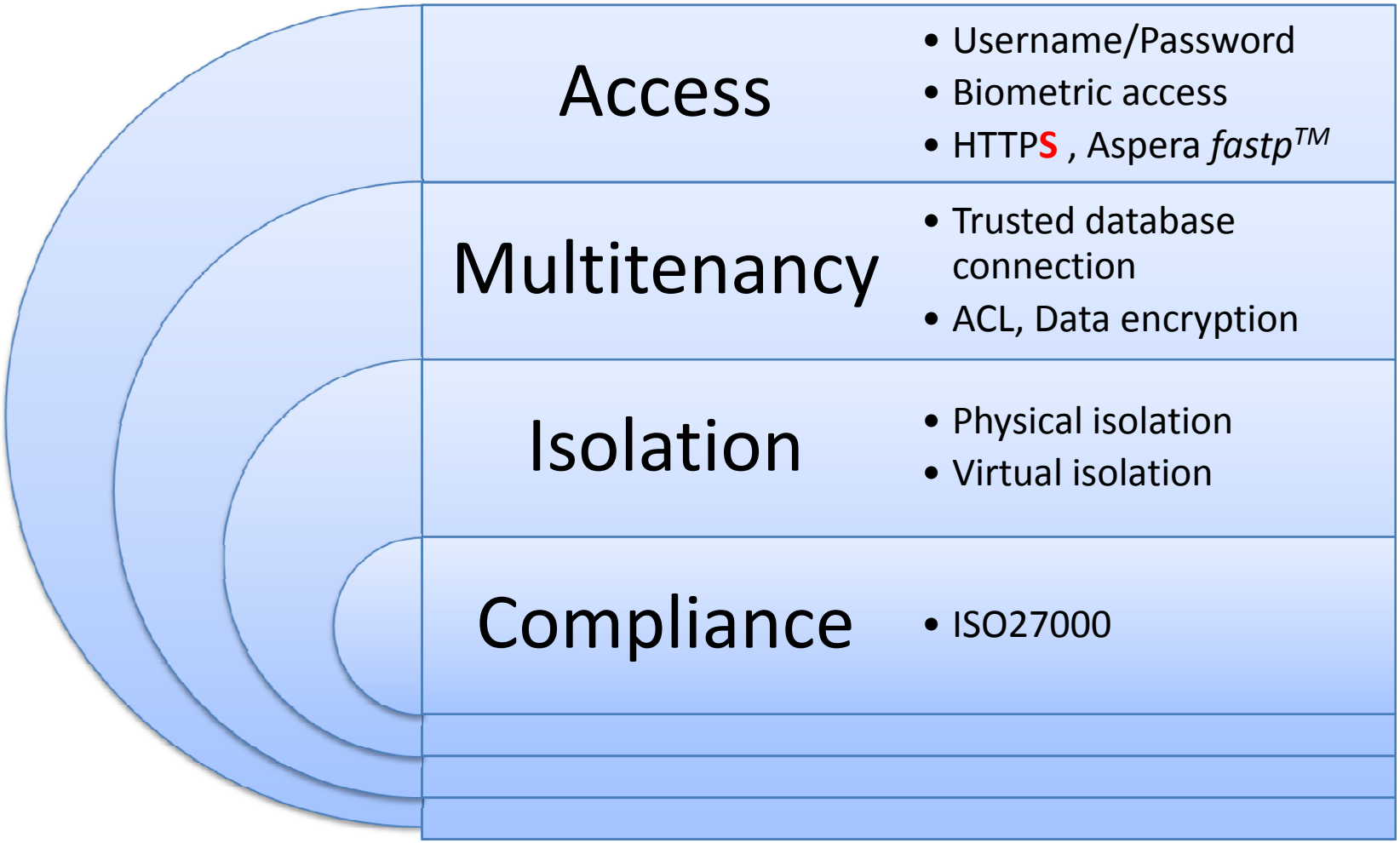
Managed
Data Structure

Managed
Task

Safe

Backup

Security

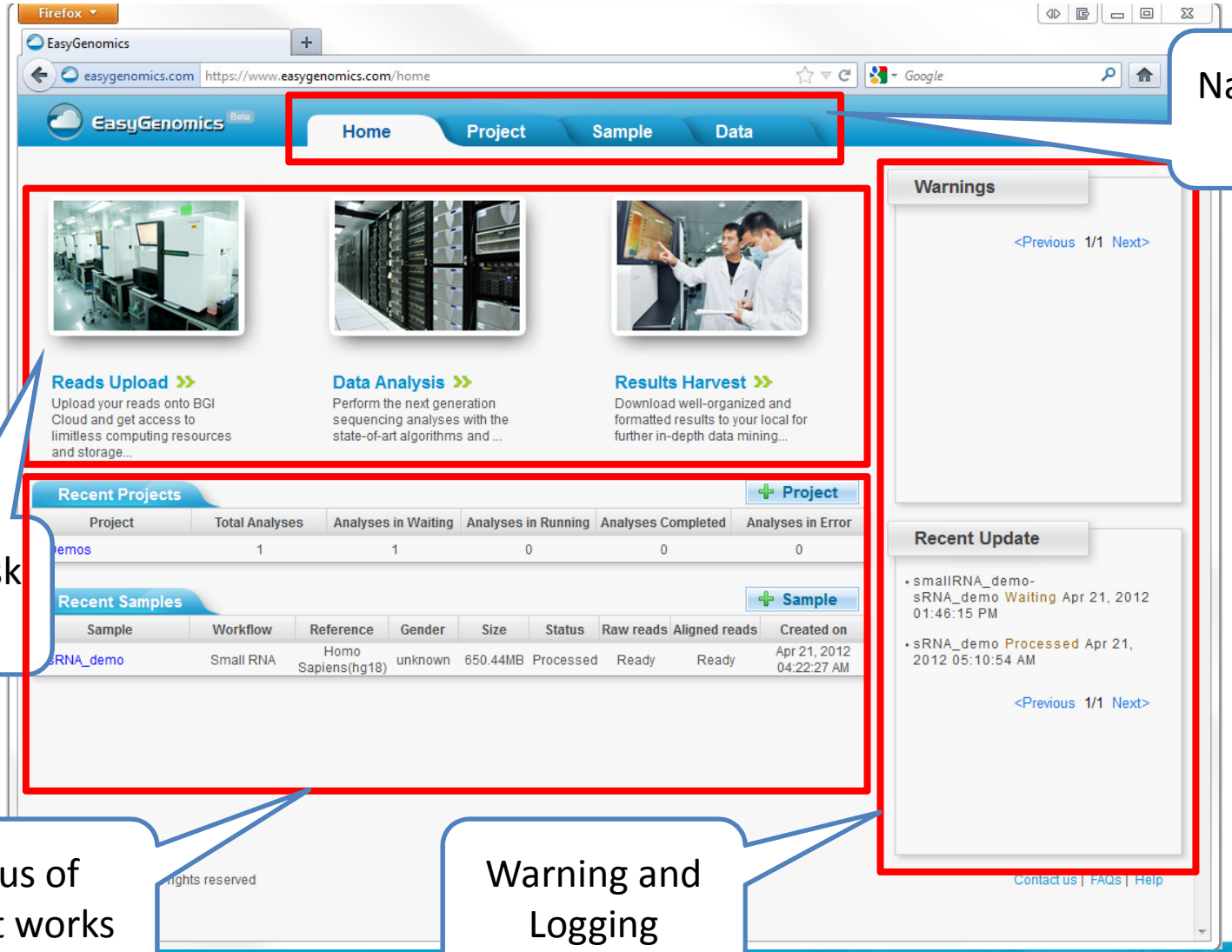

Access
• Username/Password
• Biometric access
• HTTPS , Aspera fastp™
Multitenancy
• Trusted database connection
• ACL, Data encryption
Isolation
• Physical isolation
• Virtual isolation
Compliance
• ISO27000

Introduction to EasyGenomics™

Xing Xu, Ph.D

Senior Product Manager

Homepage



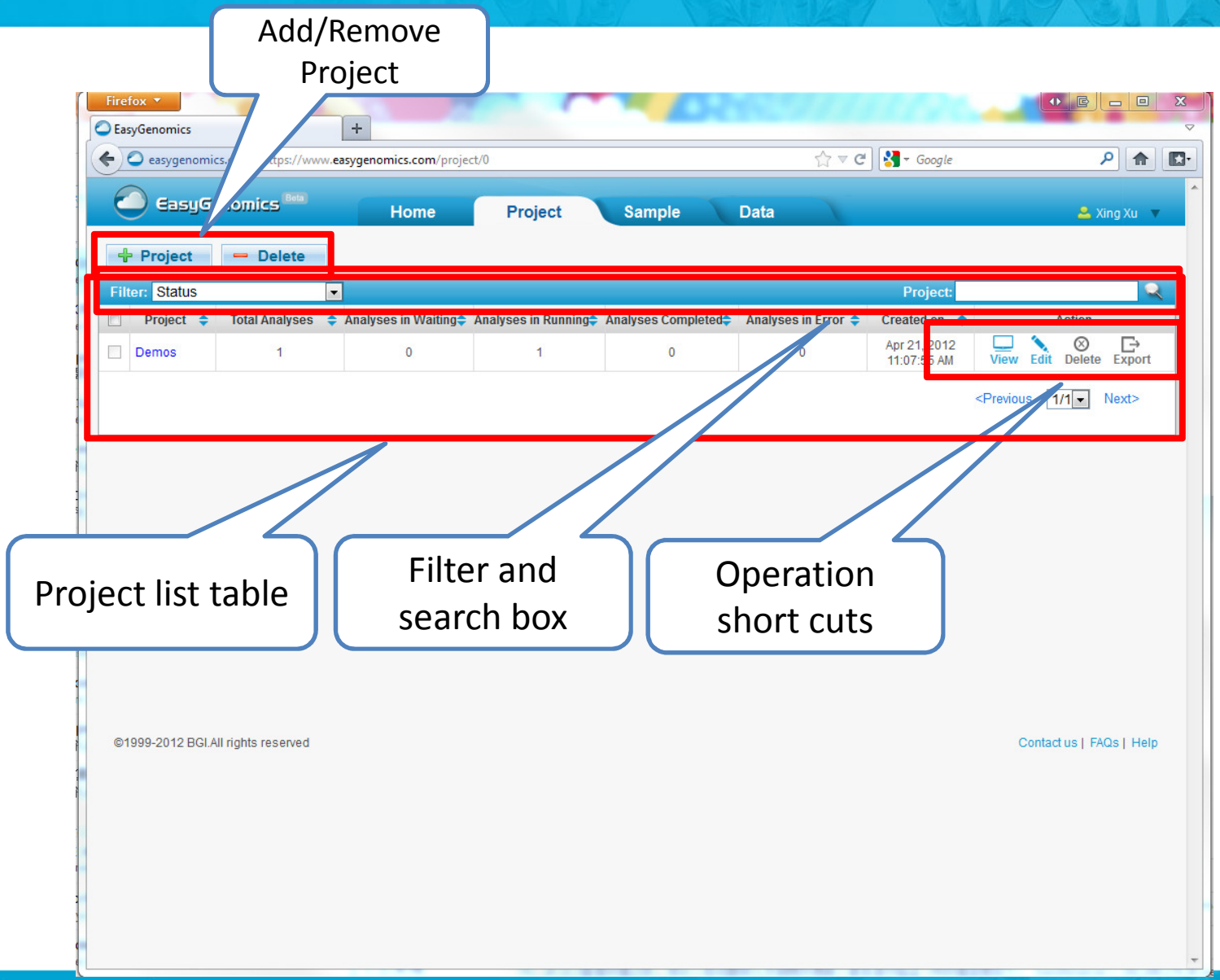
Navigation Tabs

Three task portals

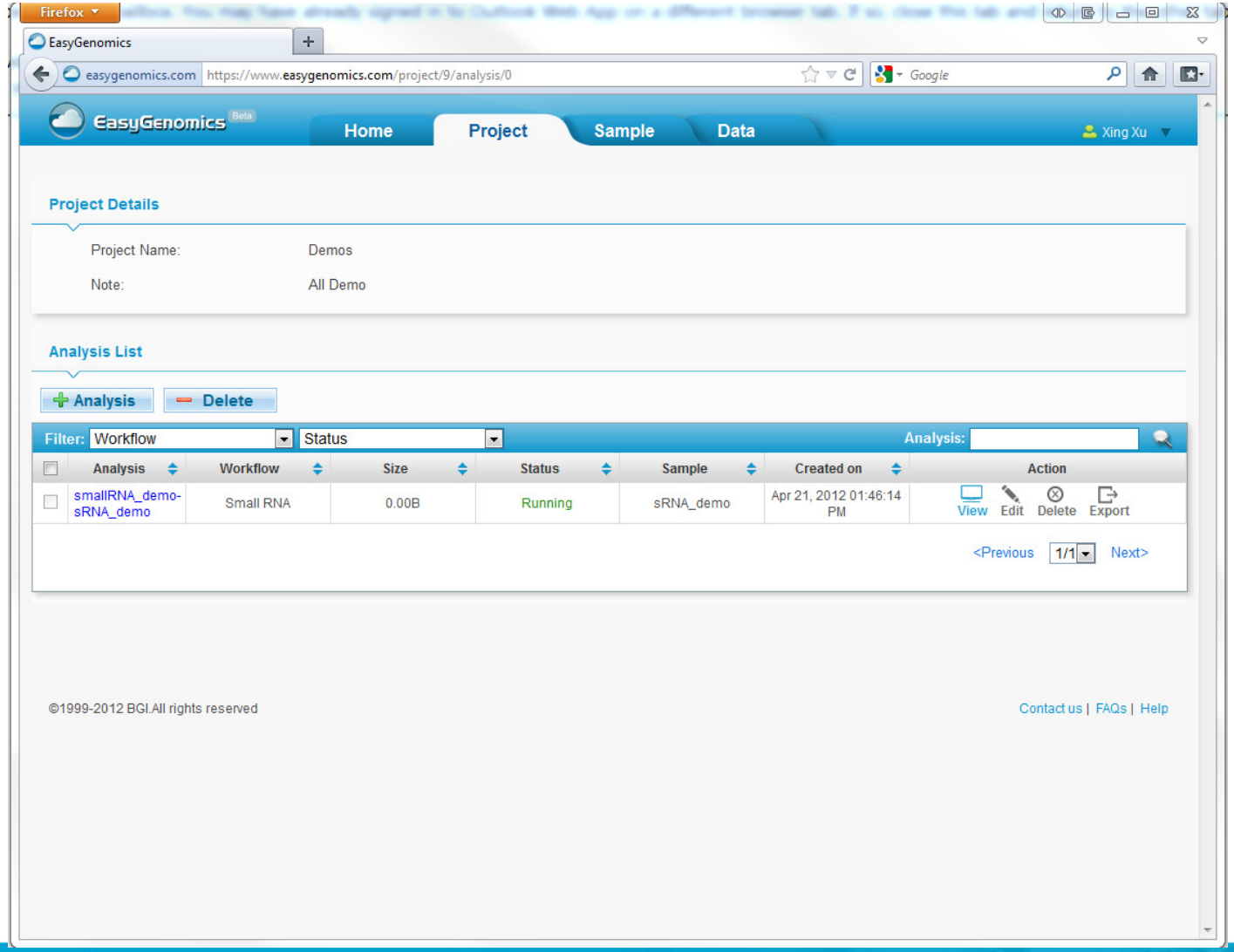
Status of recent works

Warning and Logging

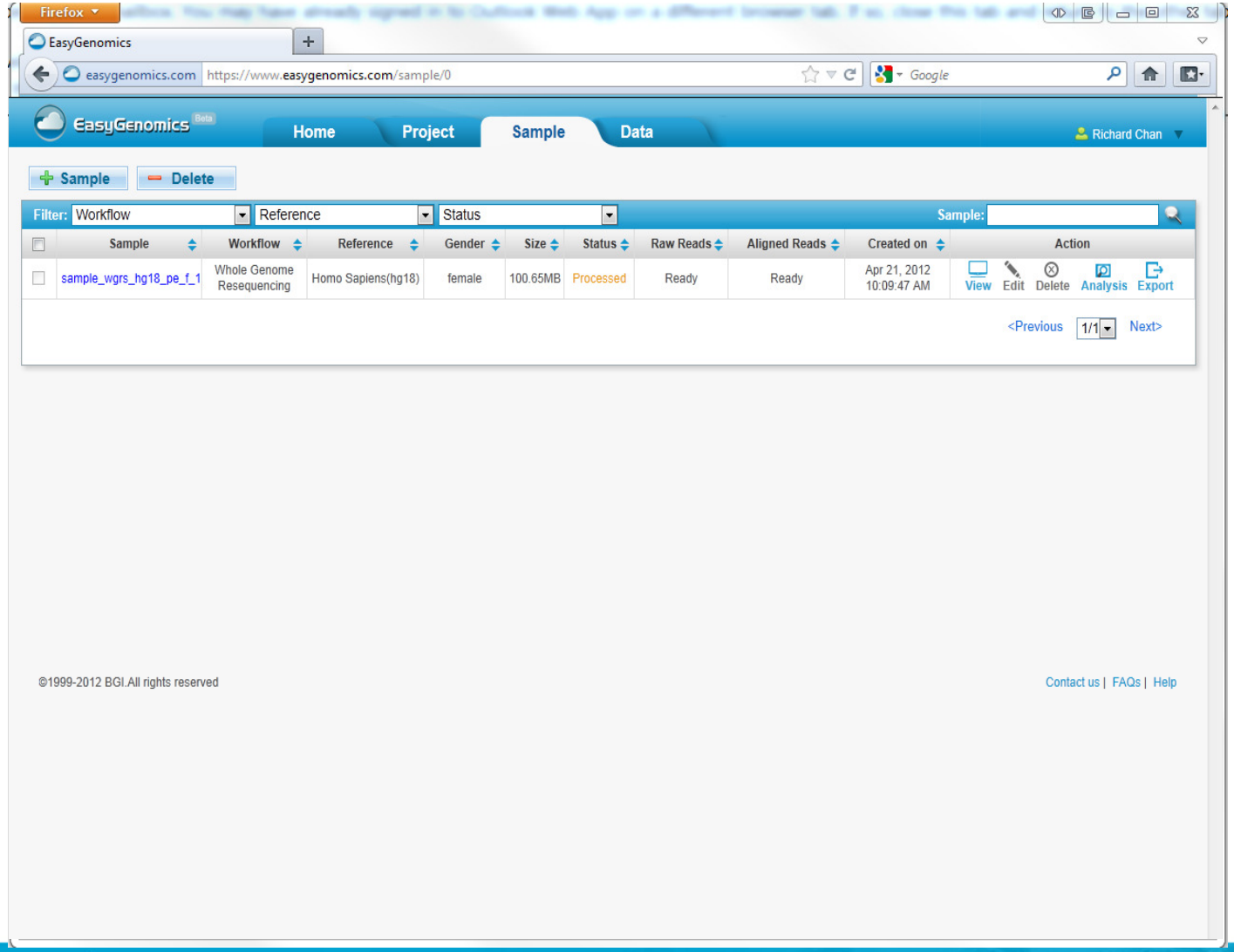
Project Table



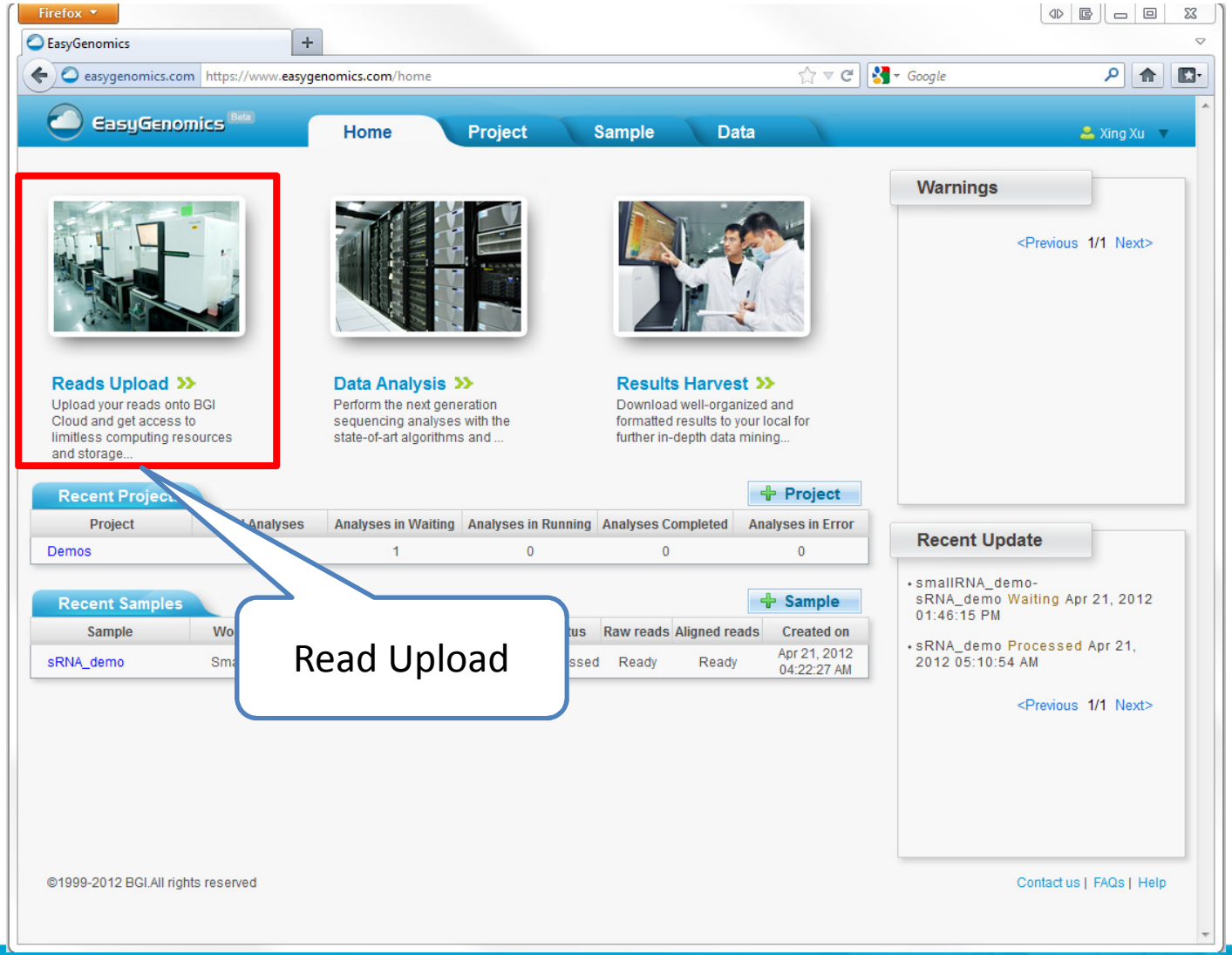
Analysis Table



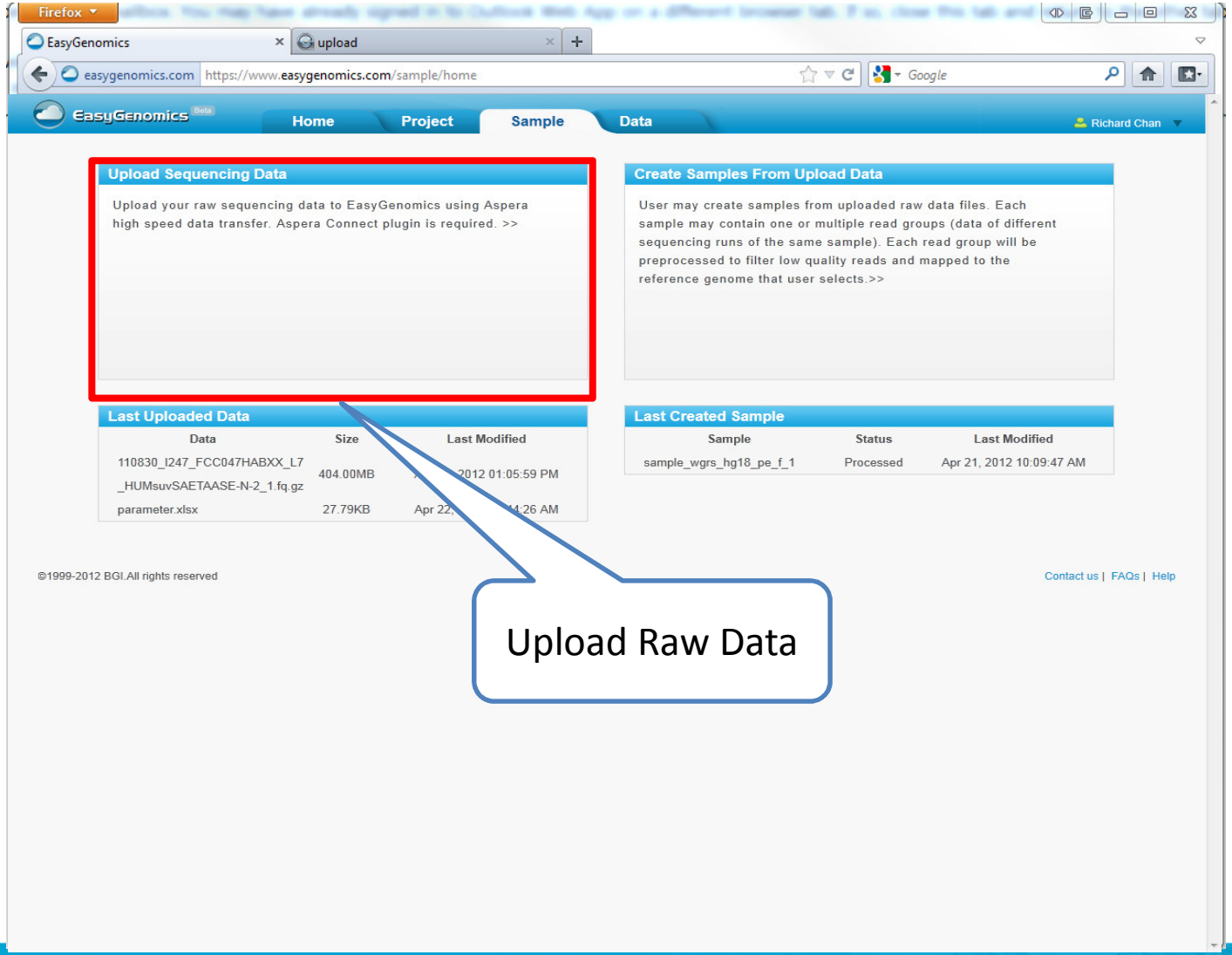
Sample Table



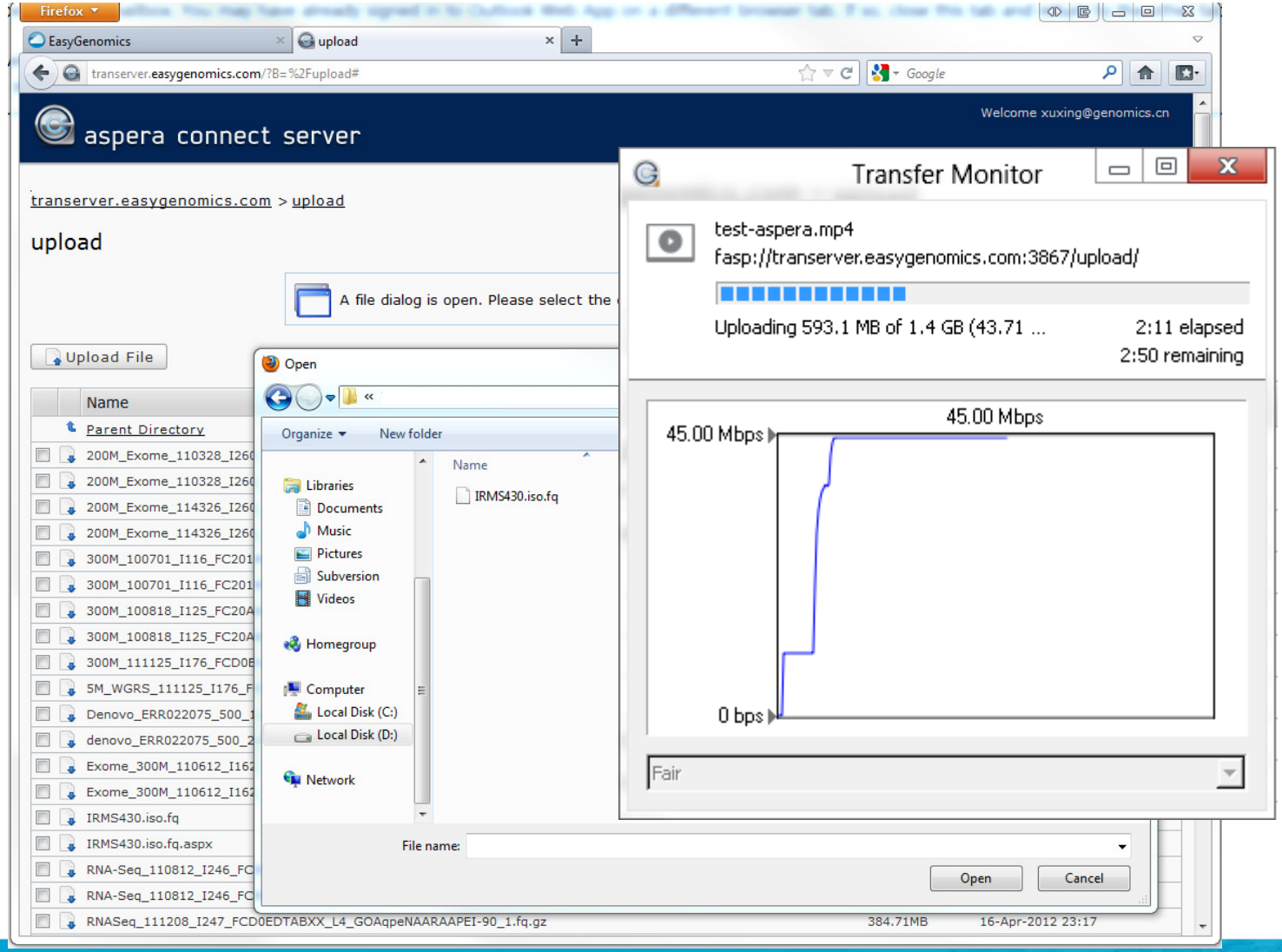
Read Upload Portal



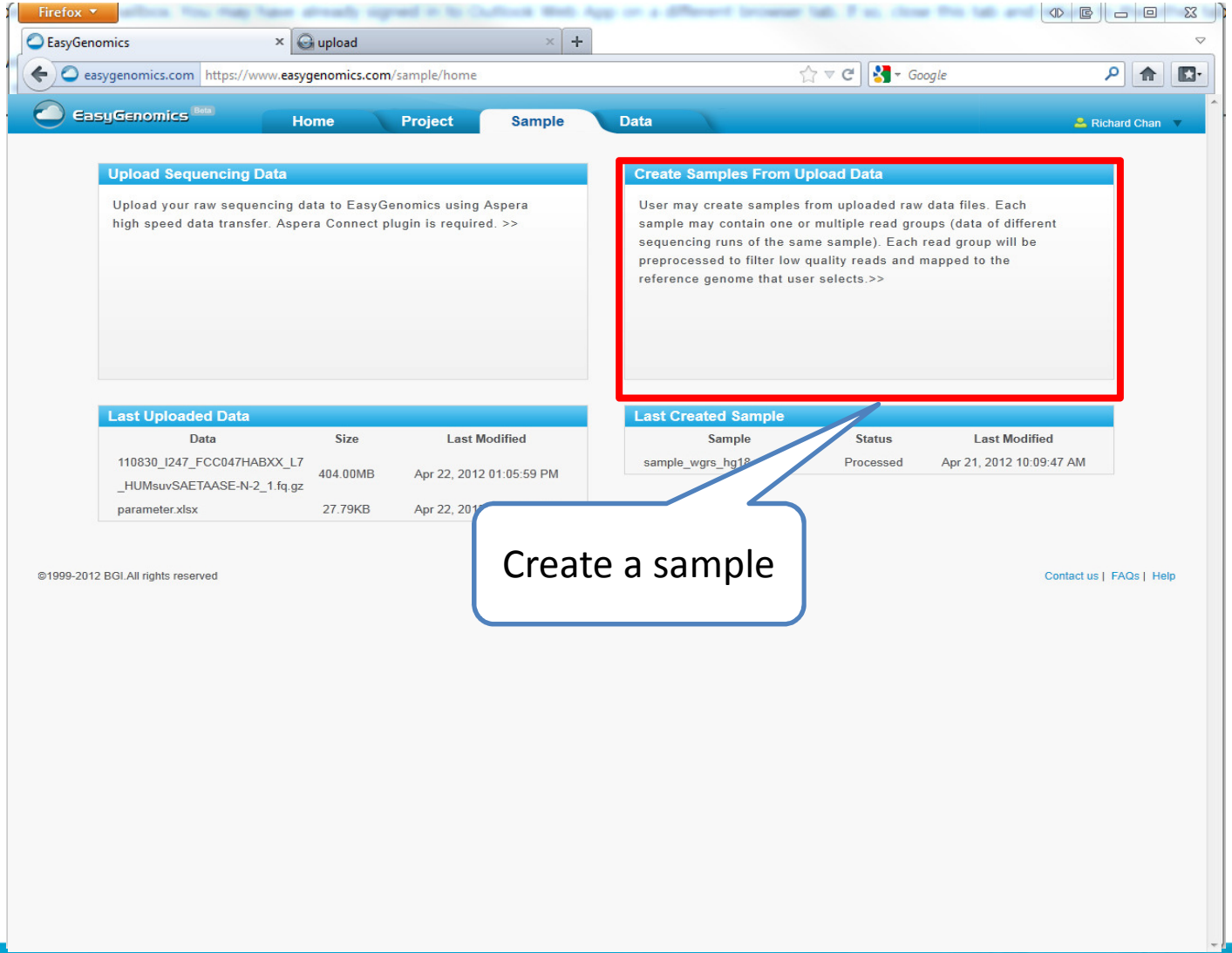
Create a Sample



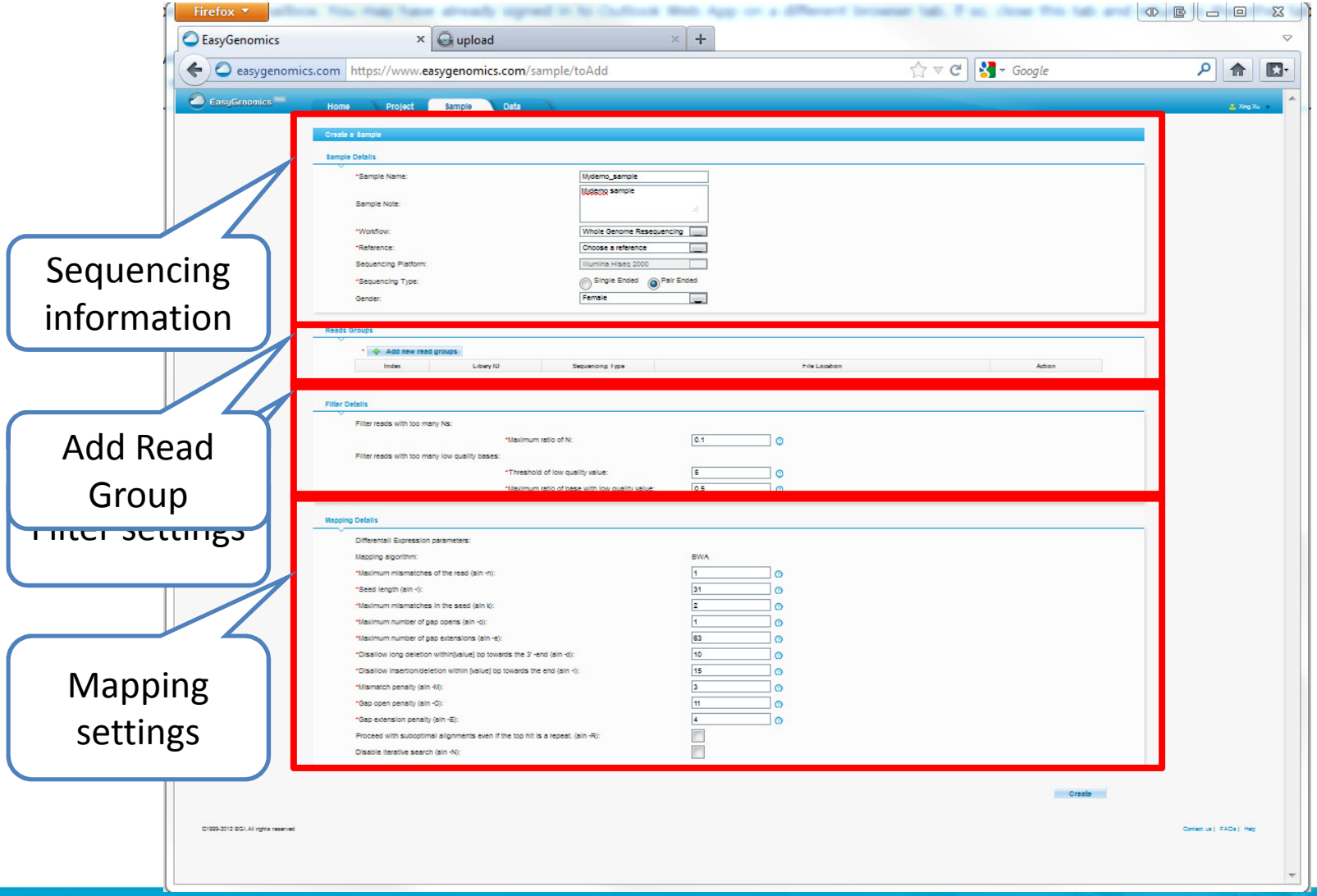
Upload Raw Reads (Aspera connect server)



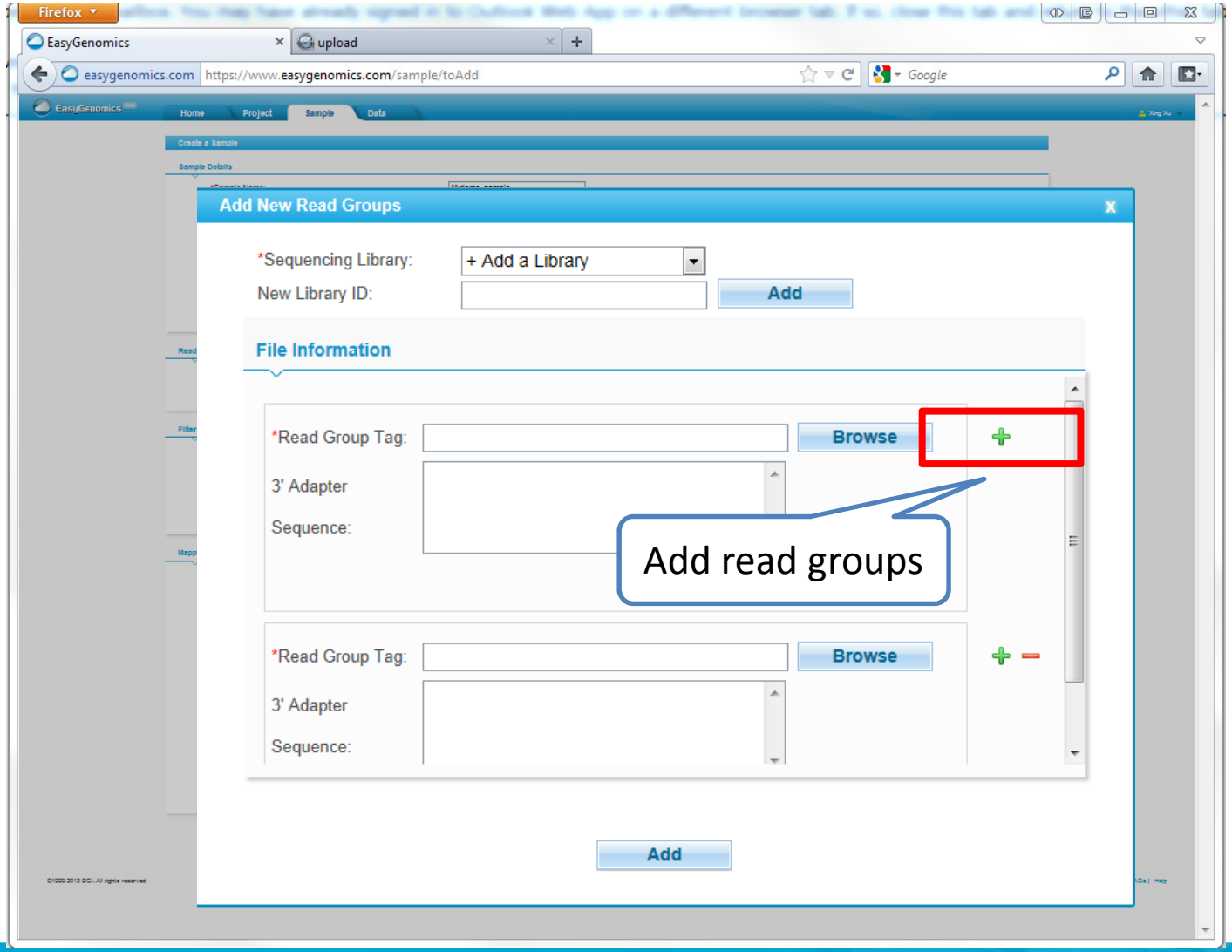
Create a Sample



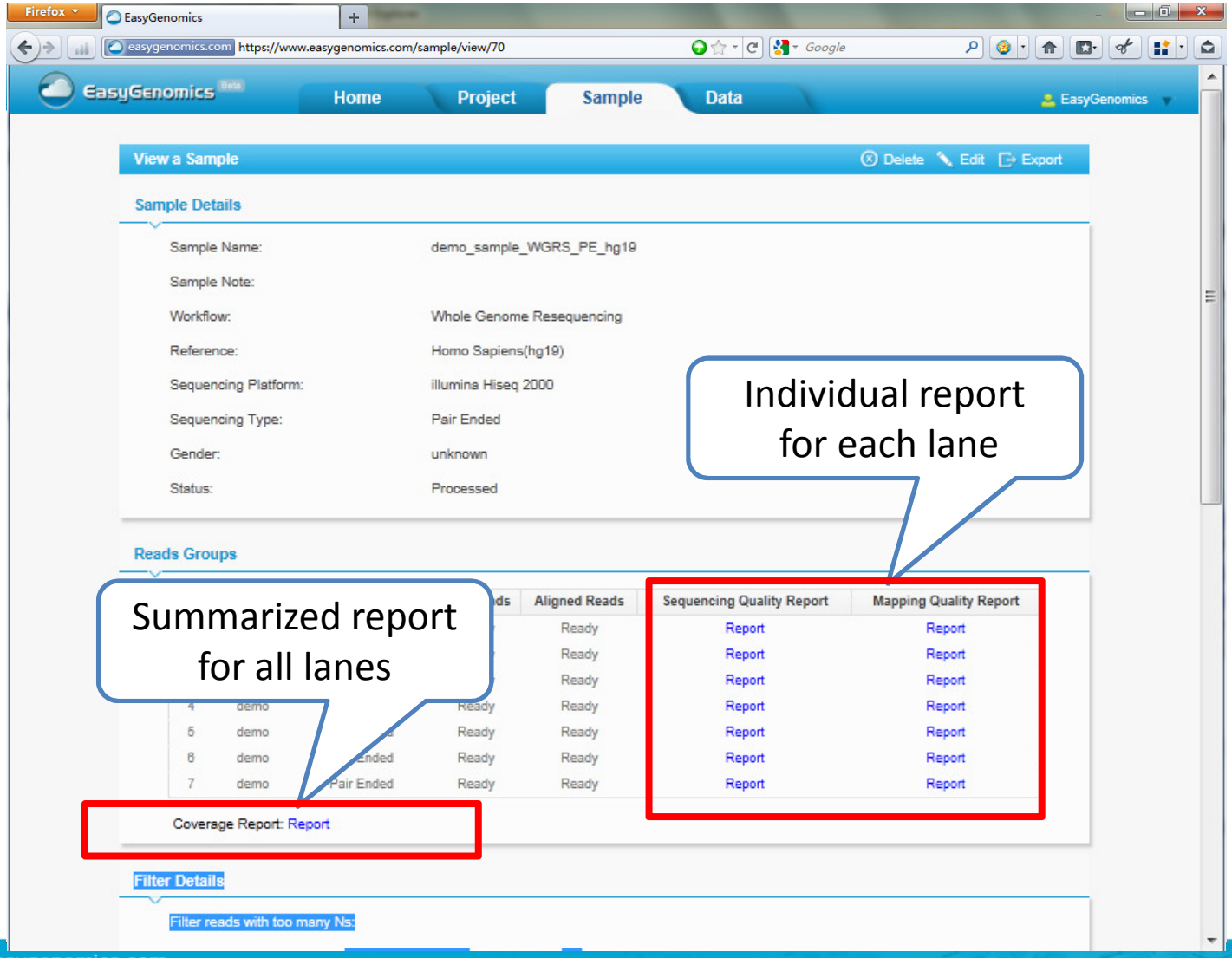
Create a Sample



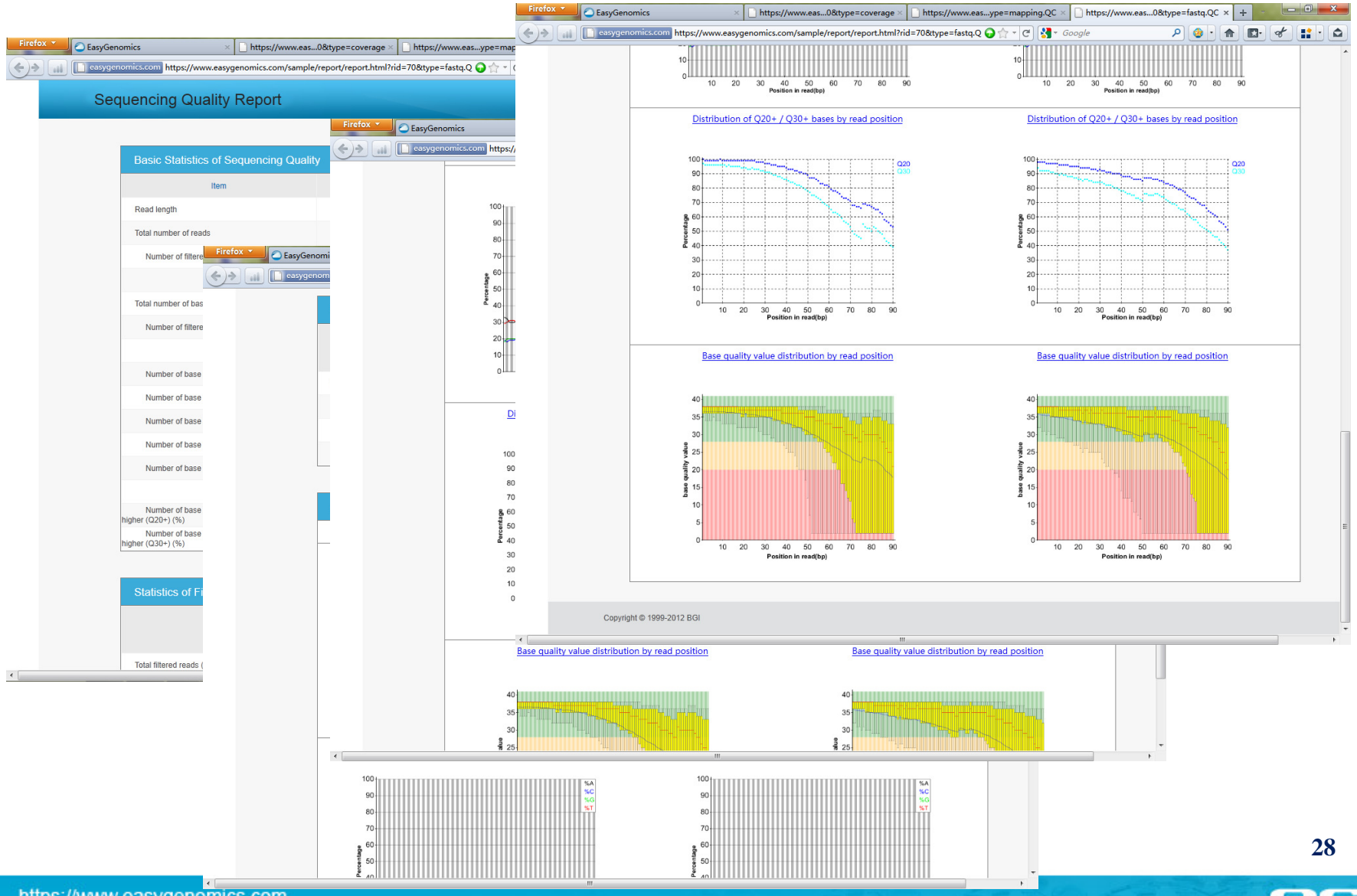
Create a Sample



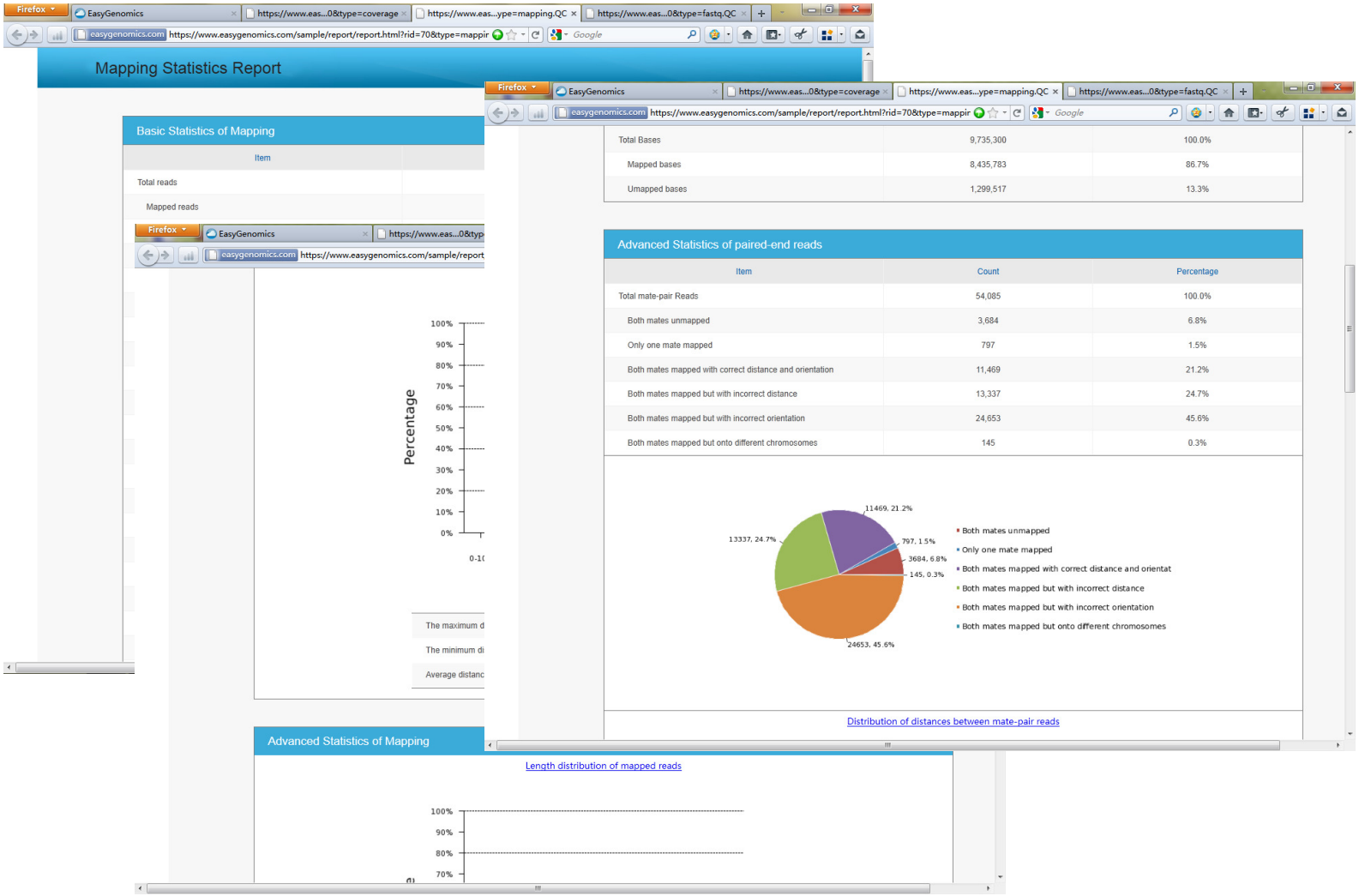
Sample Page


Firefox
EasyGenomics
https://www.easygenomics.com/sample/view/70
Home
Project
Sample
Data
View a Sample
Delete
Edit
Export
Sample Details
Sample Name: demo_sample_WGRS_PE_hg19
Sample Note:
Workflow: Whole Genome Resequencing
Reference: Homo Sapiens(hg19)
Sequencing Platform: illumina Hiseq 2000
Sequencing Type: Pair Ended
Gender: unknown
Status: Processed
Reads Groups
Aligned Reads
Sequencing Quality Report
Mapping Quality Report
Coverage Report: Report
Filter Details
Filter reads with too many Ns:
Individual report for each lane
Summarized report for all lanes

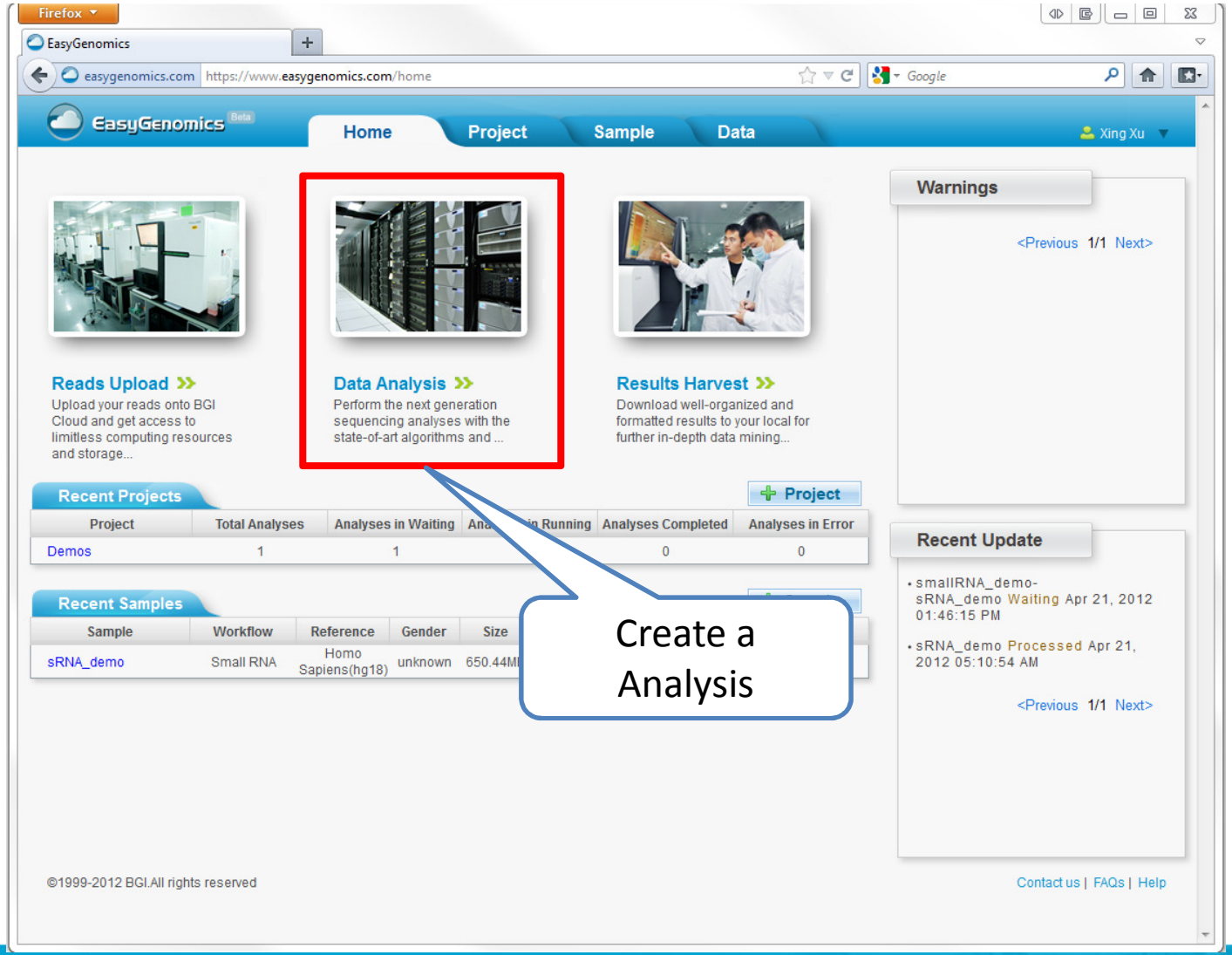
Sequencing Quality Report



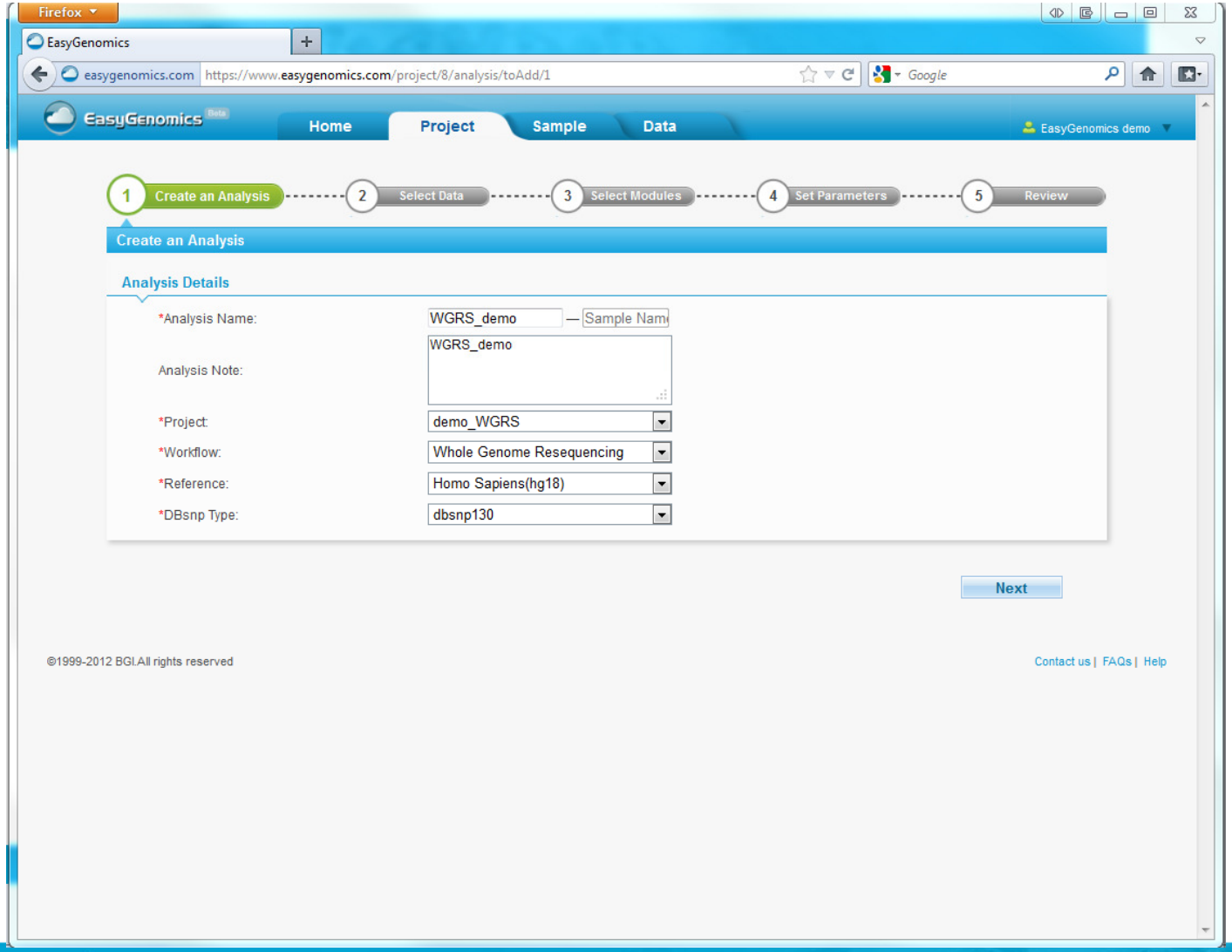
Mapping Report



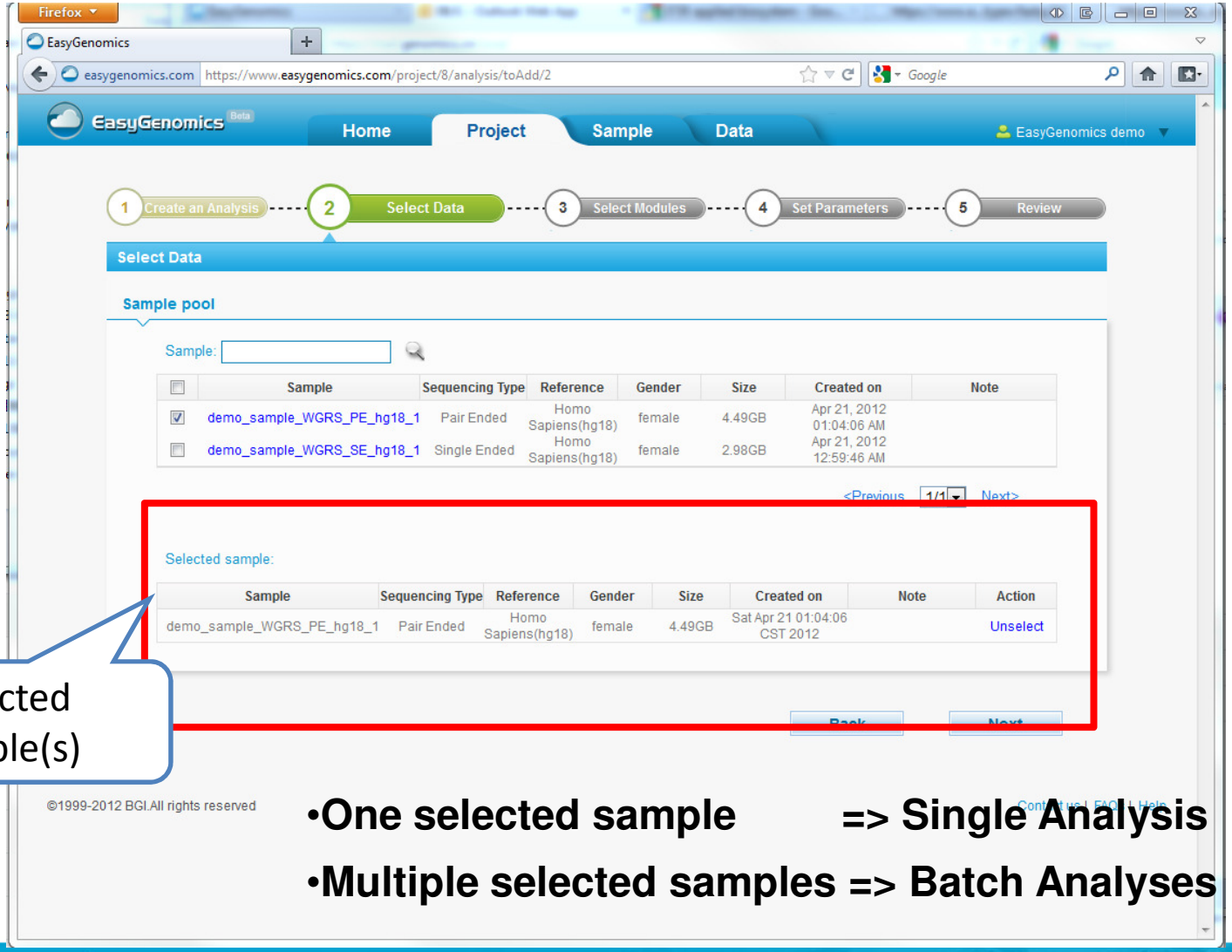
Data Analysis Portal



Create an Analysis



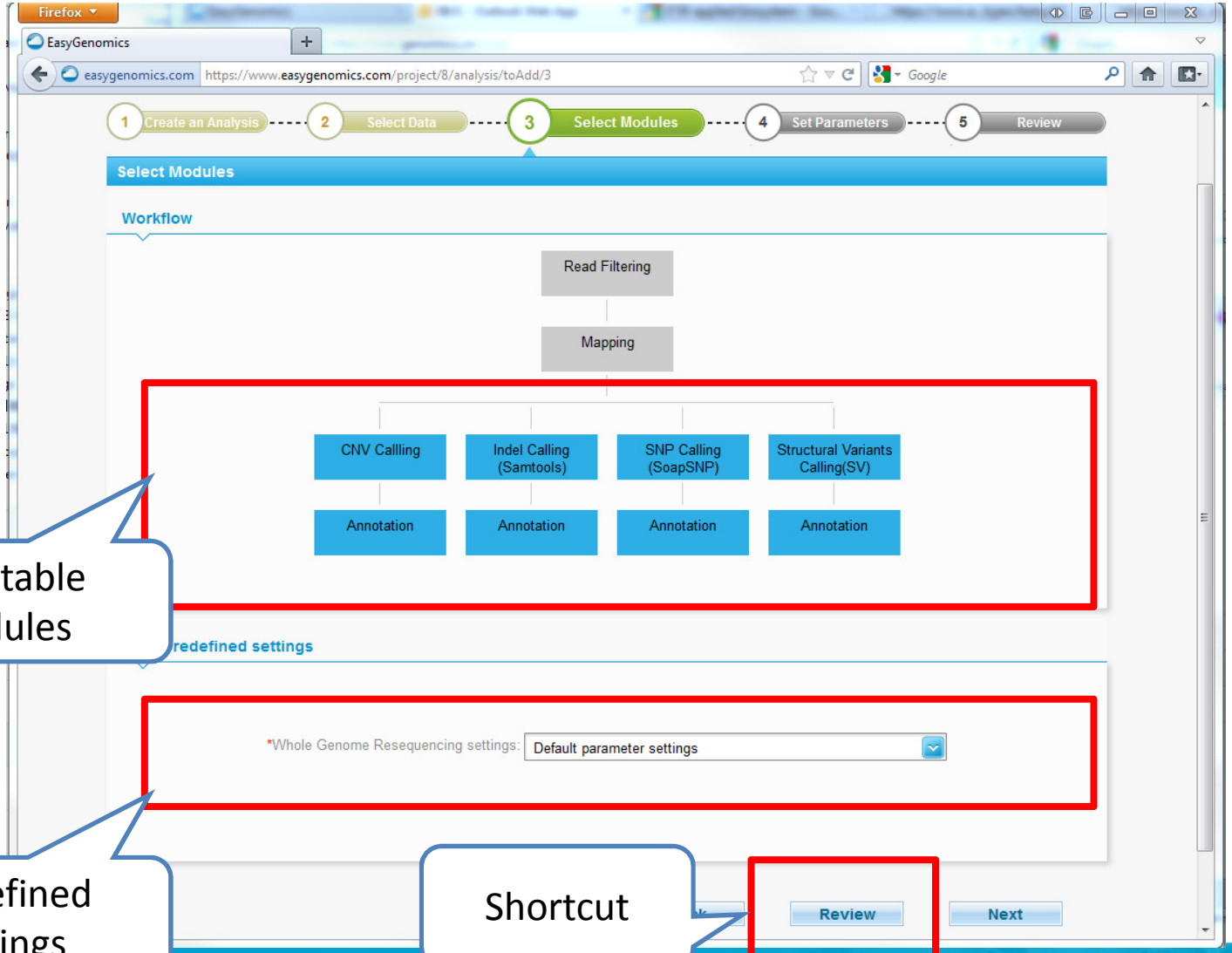
Create an Analysis



Selected sample(s)

- **One selected sample => Single Analysis**
- **Multiple selected samples => Batch Analyses**

Create an Analysis

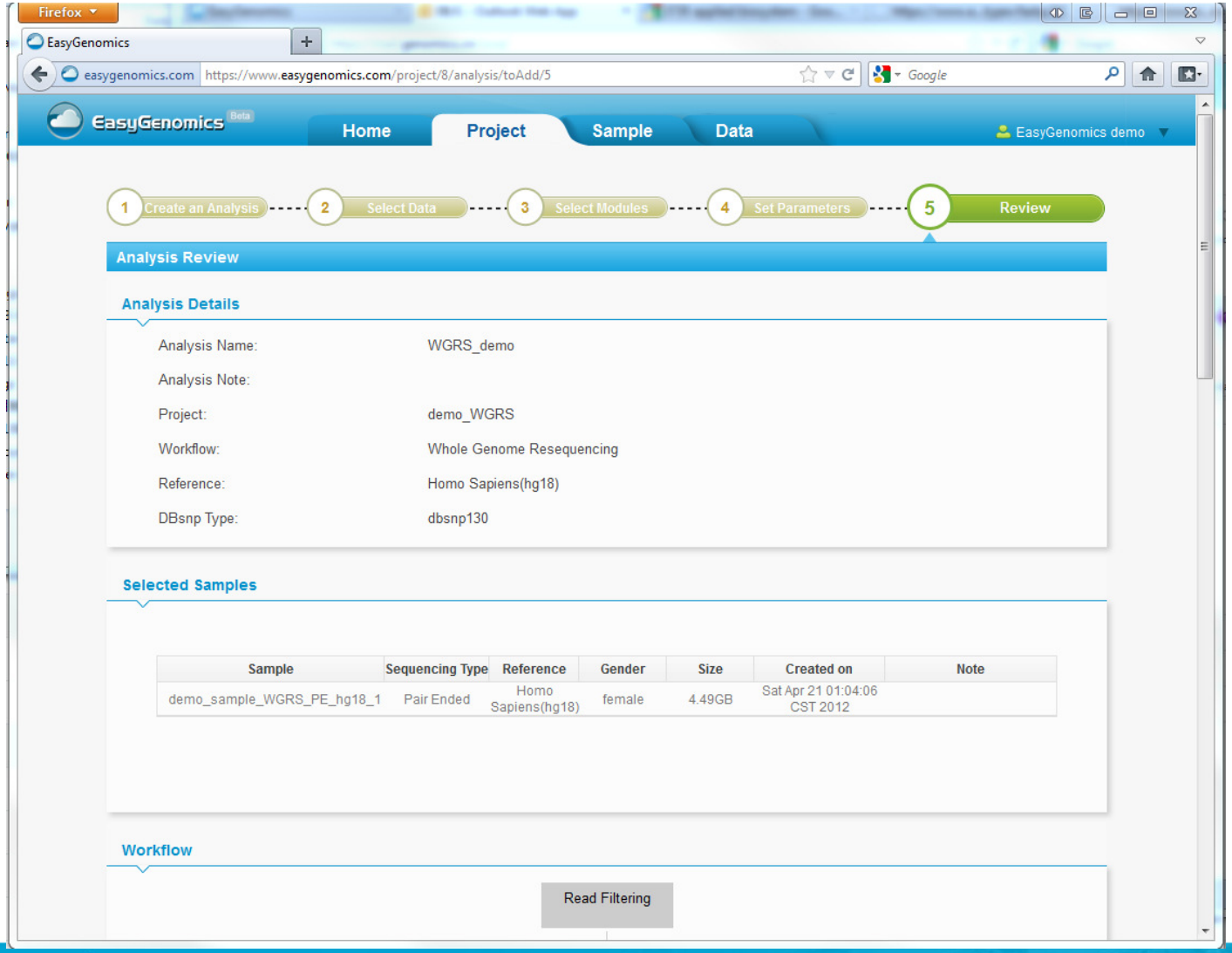


Selectable modules

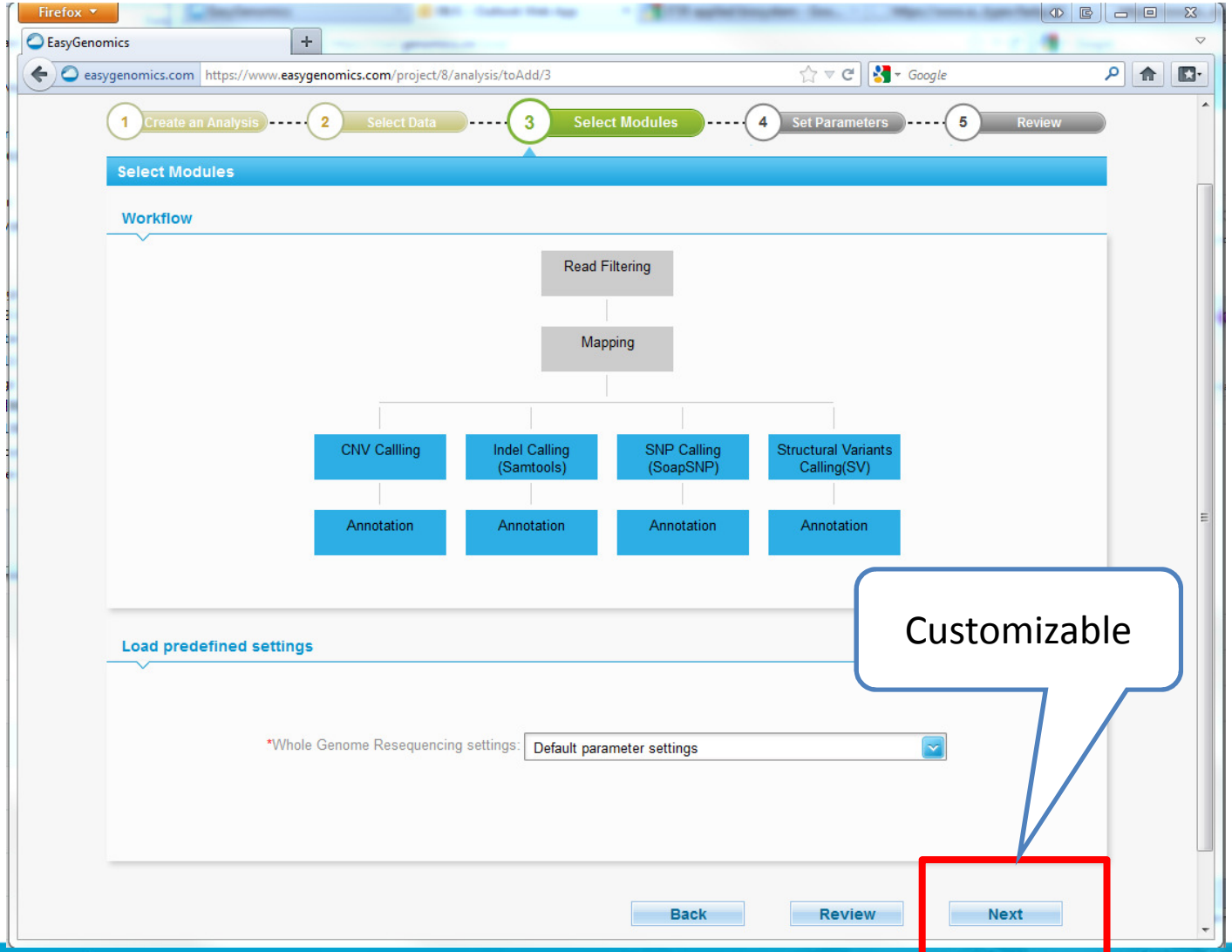
Predefined Settings

Shortcut

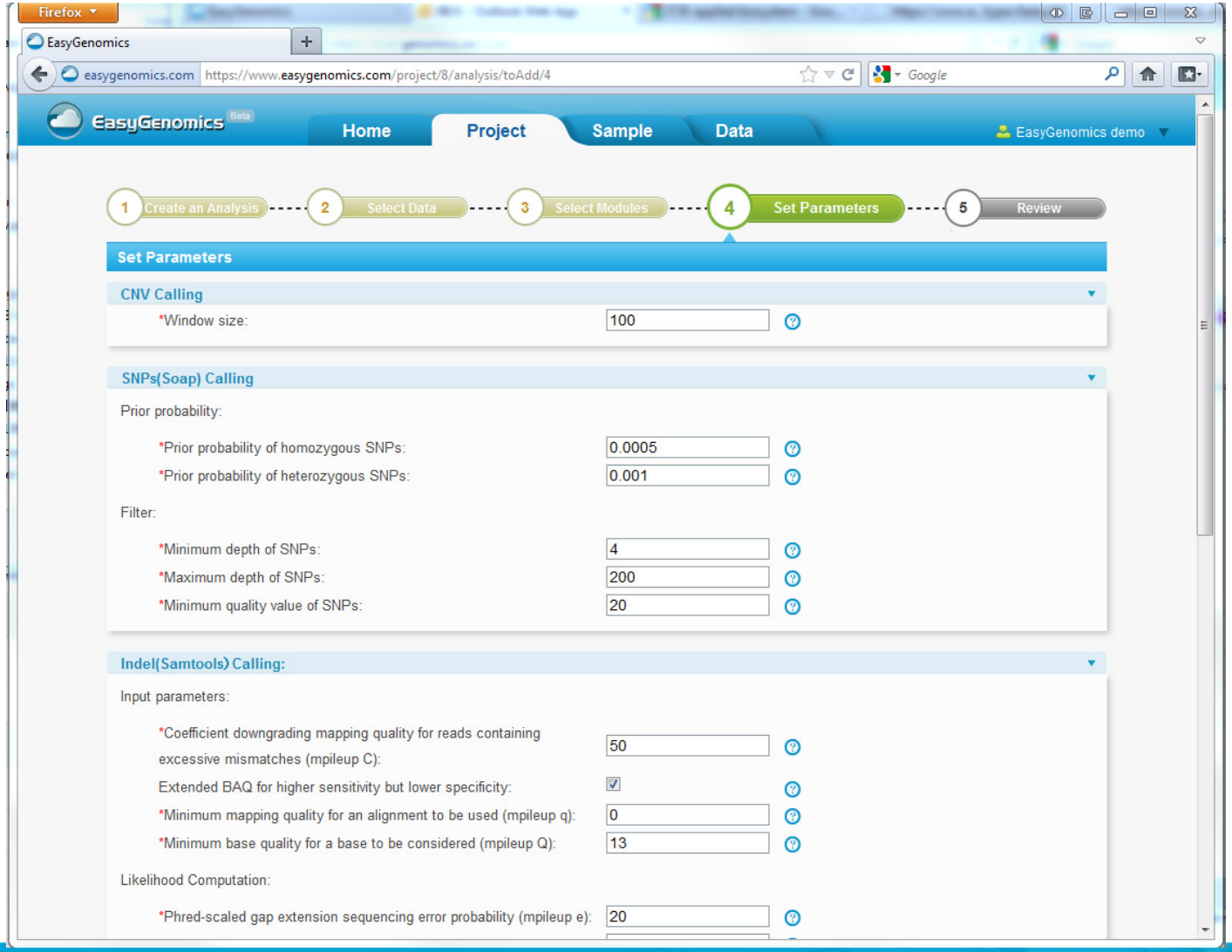
Create an Analysis



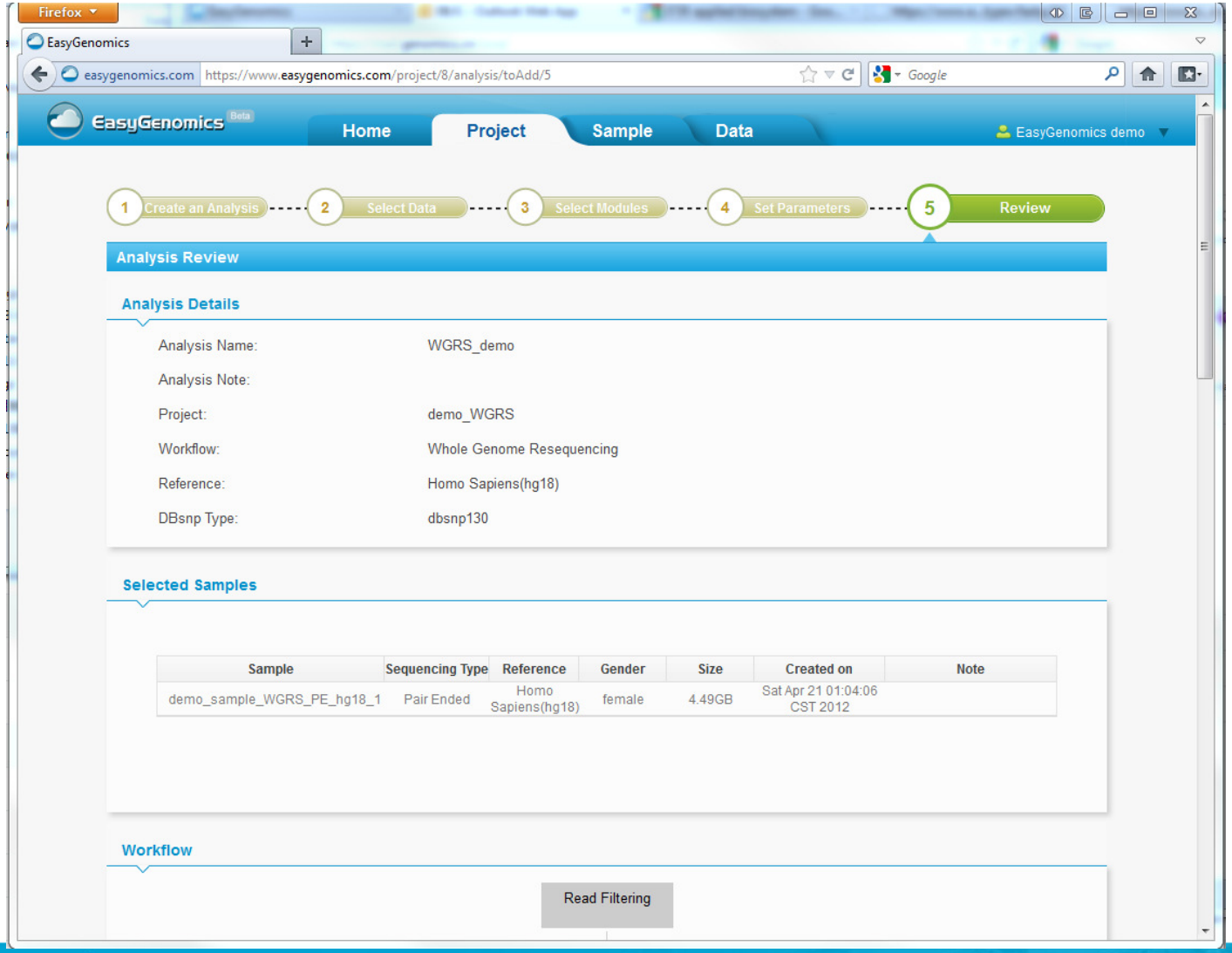
Create an Analysis



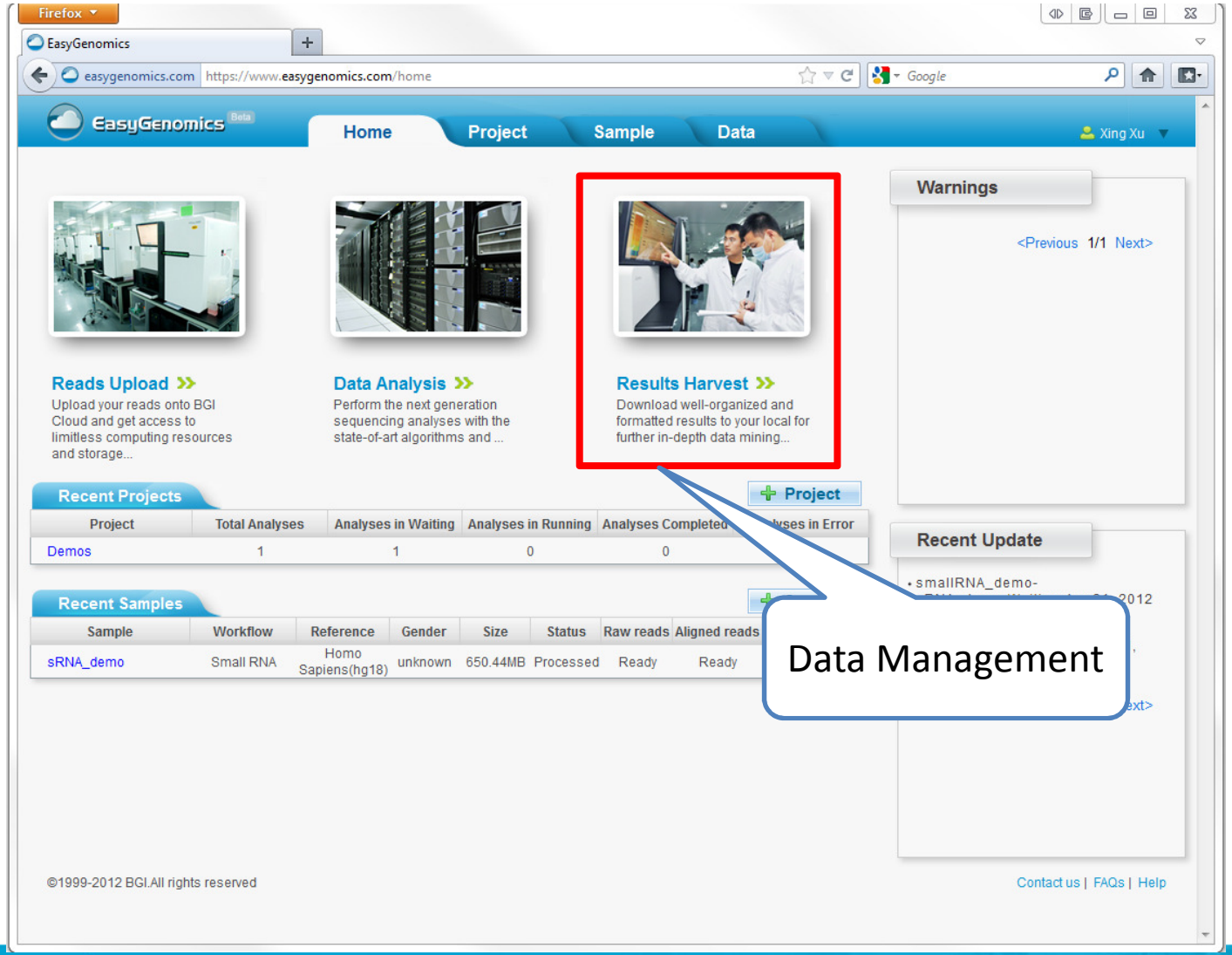
Create an Analysis



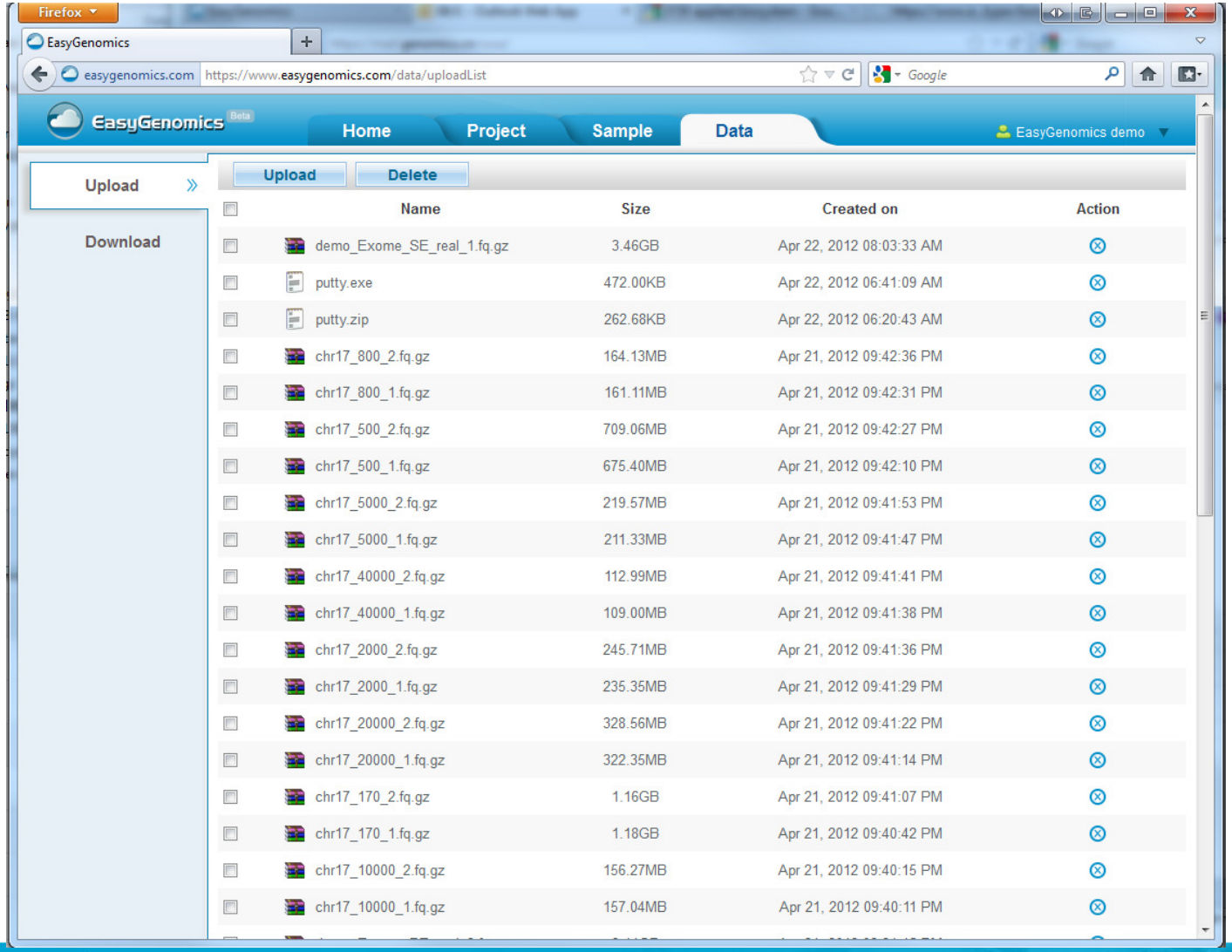
Create an Analysis



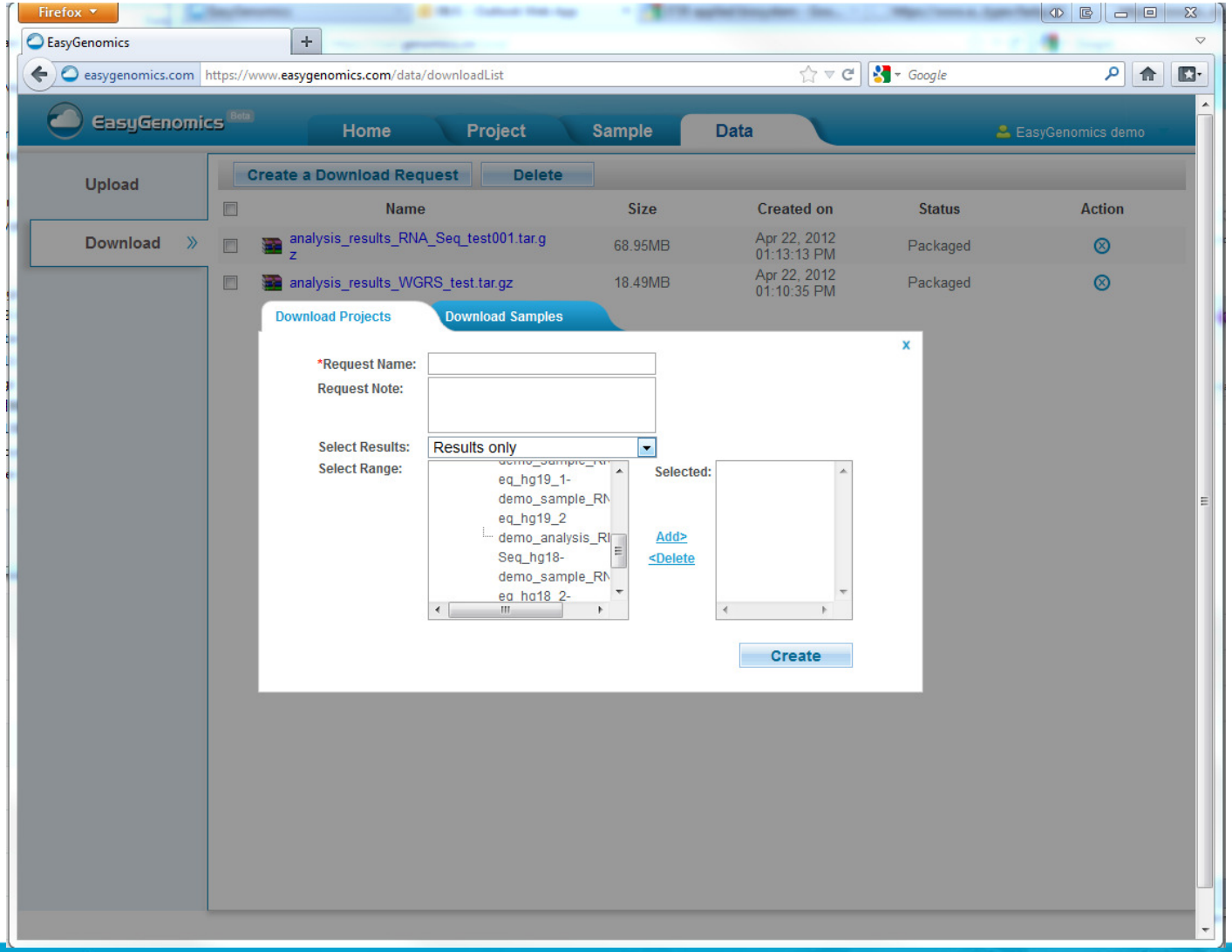
Data Harvest Portal



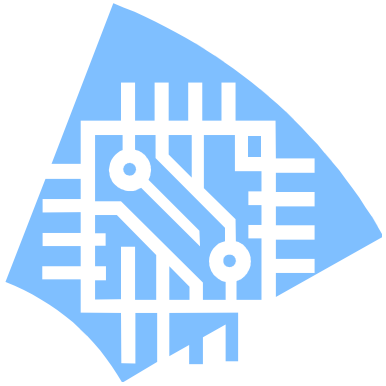
Upload Management



Download Management



LIVE DEMO



Sifei He

Director of BGI Cloud

Applications

- **Complex -Omics research**
- **Genetic testing**
- **Diagnostics**

EasyGenomics

One More Thing ☺

FREE

Ref: BOSTON

Please Visit BGI Booth @ 213

Subject to T&C

Q & A

BACKUP