High Throughput Sequencing Facility

FAPESP

1st Workshop on Multi-User-Equipment and Facilities
04/06/2014

Piotr Mieczkowski, PhD
Department of Genetics, School of Medicine at UNC

Outline of the talk

- Organization of Genomic Core in Polish Academy of Science (IBB) in 1998 (Poland)
- User perspective– Microarray Facility at University of North Carolina (USA).
- High Throughput Sequencing Facility at University of North Carolina (USA).
- Discussion

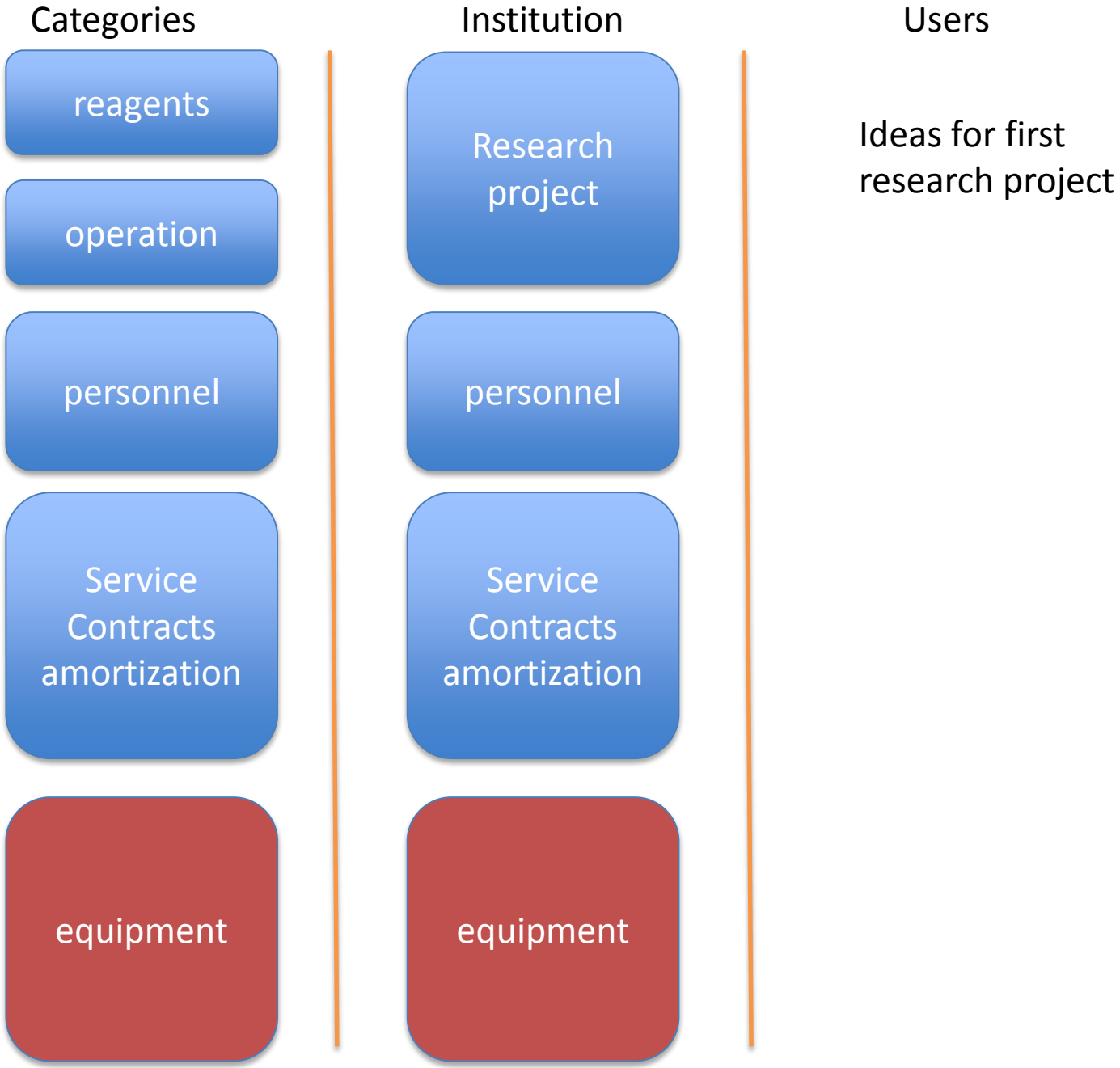
Beginning of Genomics (IBB, PL)

Institute of Biochemistry and Biophysics
Polish Academy of Sciences

Genomic Core Lab in IBB (1997)

We had machines and people dedicated to make the facility successful, but we had no funding…

Initial start up for Facility development (IBB, PL cont.)



Initial project for Genomics Lab (IBB-Poland) in 1997

- We are to small to compete with Europe or US institutions
- We do not have capacity and funding to sequence genomes
- We need to find our niche which we can develop in the future
- We need to increase our visibility and collaboration with other Polish research institutes.

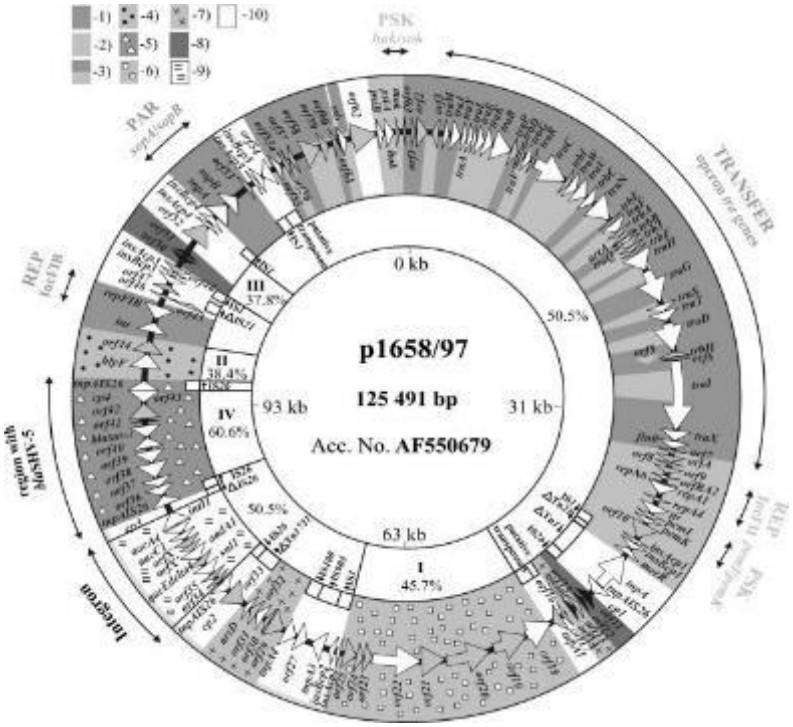
Sequencing plasmids will be the best option for small sequencing Genomics Lab

ANTIMICROB. AGENTS AND CHEMOTHERAPY, Apr. 2007, p. 1167–1171 Vol. 51, No. 4
0066-4804/07/$08.00+0 doi:10.1128/AAC.00772-06

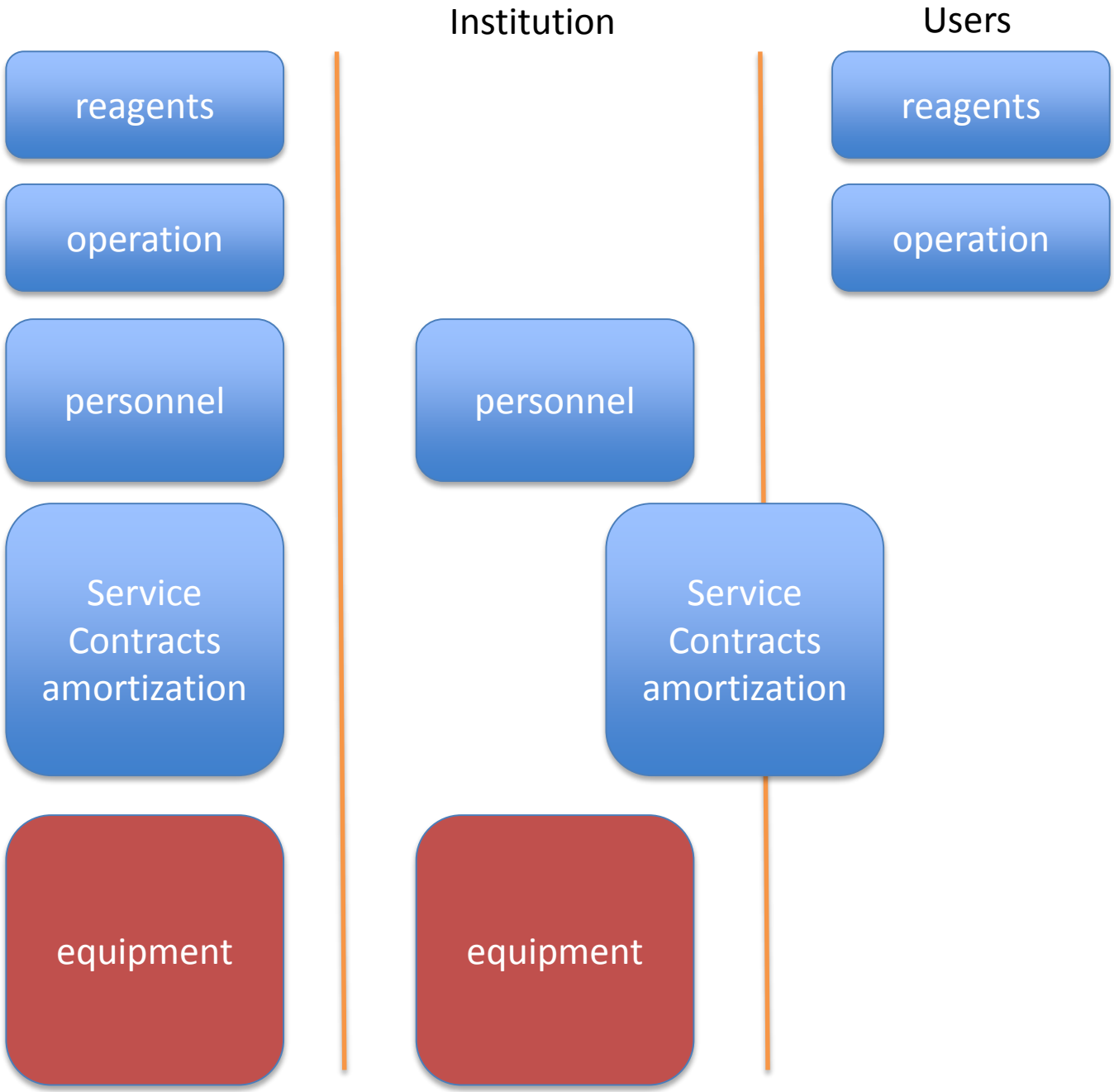
Mosaic Structure of p1658/97, a 125-Kilobase Plasmid Harboring an Active Amplicon with the Extended-Spectrum β-Lactamase Gene bla_{SHV-5}

M. Zienkiewicz,[1]* I. Kern-Zdanowicz,[1] M. Gołębiewski,[1] J. Żyłtfiska,[1] P. Mieczkuwski,[1]‡ M. Gniadkowski,[2] J. Bardowski,[1] and P. Cegłowski[1]

Department of Microbial Biochemistry, Institute of Biochemistry and Biophysics of Polish Academy of Sciences,[1] and National Institute of Public Health,[2] Warsaw, Poland



Structure of Facility funding



Conclusions

- Institutional support is necessary for successful start up of new facility.
- Institutional investment in one relevant project was a good start for facility and brought attention of all investigators.
- Federal institution supporting many aspects of operation in facility (salaries, amortization of equipment) gives opportunity for better pricing even if the cost of the reagents is high (higher than in US).
- The same pricing for service provided was offered to all customers, both domestic and foreign.


P22

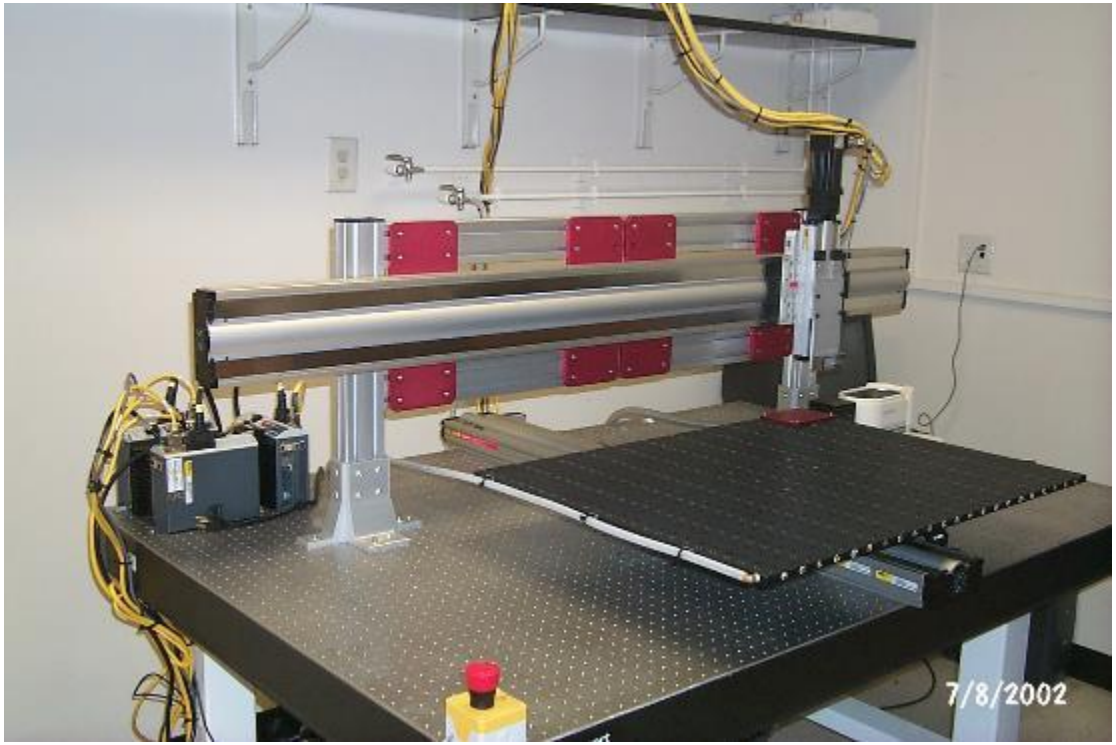

Canada
United States
Mexico
Olympia
Washington
Salem
Eugene
Oregon
Boise
Idaho
Helena
Montana
Bismarck
North Dakota
Minnesota
Pierre
South Dakota
St. Paul
Wisconsin
Wyoming
Cheyenne
Nebraska
Madison
Iowa
Des Moines
Lincoln
Denver
Colorado
Topeka
Kansas
Jefferson City
Missouri
Springfield
Illinois
Indianapolis
Indiana
Lansing
Michigan
Ottawa
Montpelier
Augusta
Maine
Albany
New York
Concord
Boston
Providence
Hartford
Trenton
New Jersey
Harrisburg
Pennsylvania
Columbus
Ohio
Charleston
West Virginia
Frankfort
Kentucky
Richmond
Virginia
Washington D.C.
Annapolis
Dover
Delaware
Raleigh
North Carolina
Nashville
Tennessee
Atlanta
Georgia
Alabama
Montgomery
Columbia
South Carolina
Tallahassee
Florida
Jackson
Mississippi
Louisiana
Baton Rouge
Little Rock
Arkansas
Oklahoma City
Oklahoma
Texas
Austin
Rio Grande R.
Santa Fe
New Mexico
Phoenix
Arizona
Santa Barbara
California
Sacramento
Carson City
Nevada
Salt Lake City
Utah
Colorado R.
Missouri R.
Columbia R.
Snake R.
Lake Winnipeg
Lake Superior
Lake Michigan
Lake Huron
Lake Erie
Lake Ontario
Ohio R.
Mississippi R.
Pacific Ocean
Atlantic Ocean
Gulf of Mexico
The Bahamas
Havana
Cuba
Hawaii
Honolulu
Alaska
Juneau
Canada
Bering Sea
0 100 mi
0 300 mi
0 800 km
0 500 mi
©1994 MAGELLAN Geographix℠ Santa Barbara, CA (800) 929-4MAP

UNC

Microarray printing

- Microarray was emerging technology in 1995-2004.
- Availability of the commercial microarray slides was limited by price. Only very rich labs could afford to buy them.
- UNC established Microarray facility equipped in medium size printer and scanner.
- User had to provide DNA or oligonucleotides for printing and pay $50 per slide for service and glass slide processing (commercial glass slide for printing -$15)

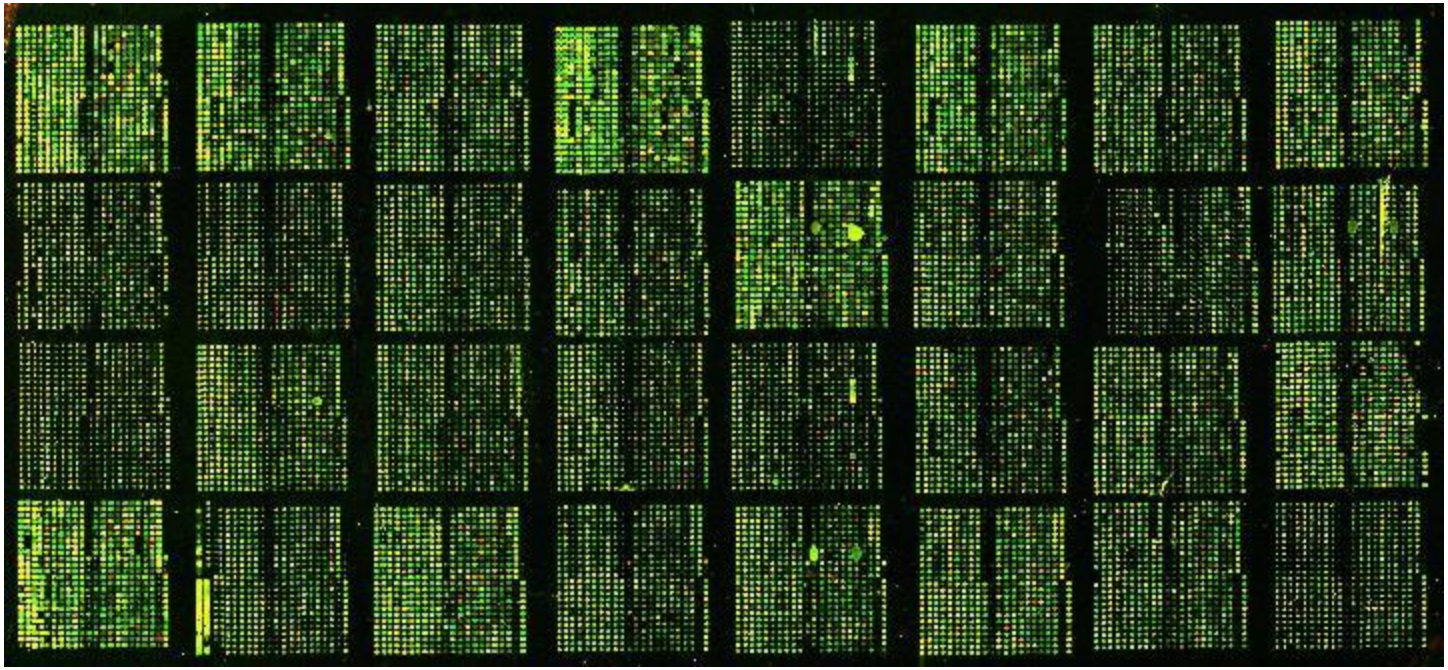
We built our own microarray printer to reduce the cost for our experiments



Array Platform

The array platform we used was based on PCR product (containing the Open Reading Frames (or ORF) and Intergenic regions) printed on the poly-lysine slides.

Below: image of the home-printed microarray slide covered by entire *Saccharomyces cerevisiae* genome.



About 12,100 spots cover the genome at ~1000 bp resolution

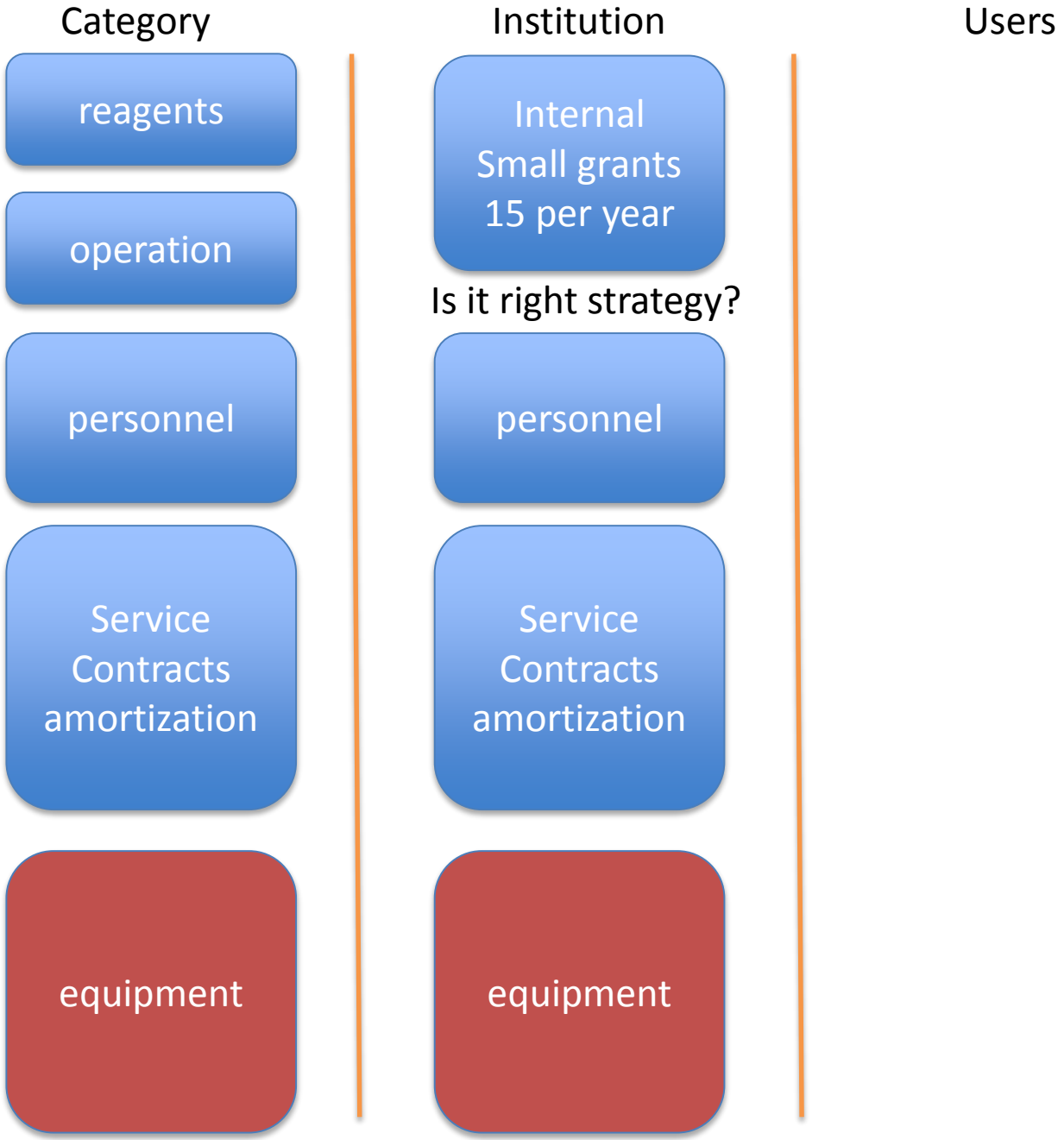
Conclusions

- Facility did not provide expected expertise in production of custom microarrays.
- It was more cost effective to prepare amplicons and build own printer than to use the facility service.

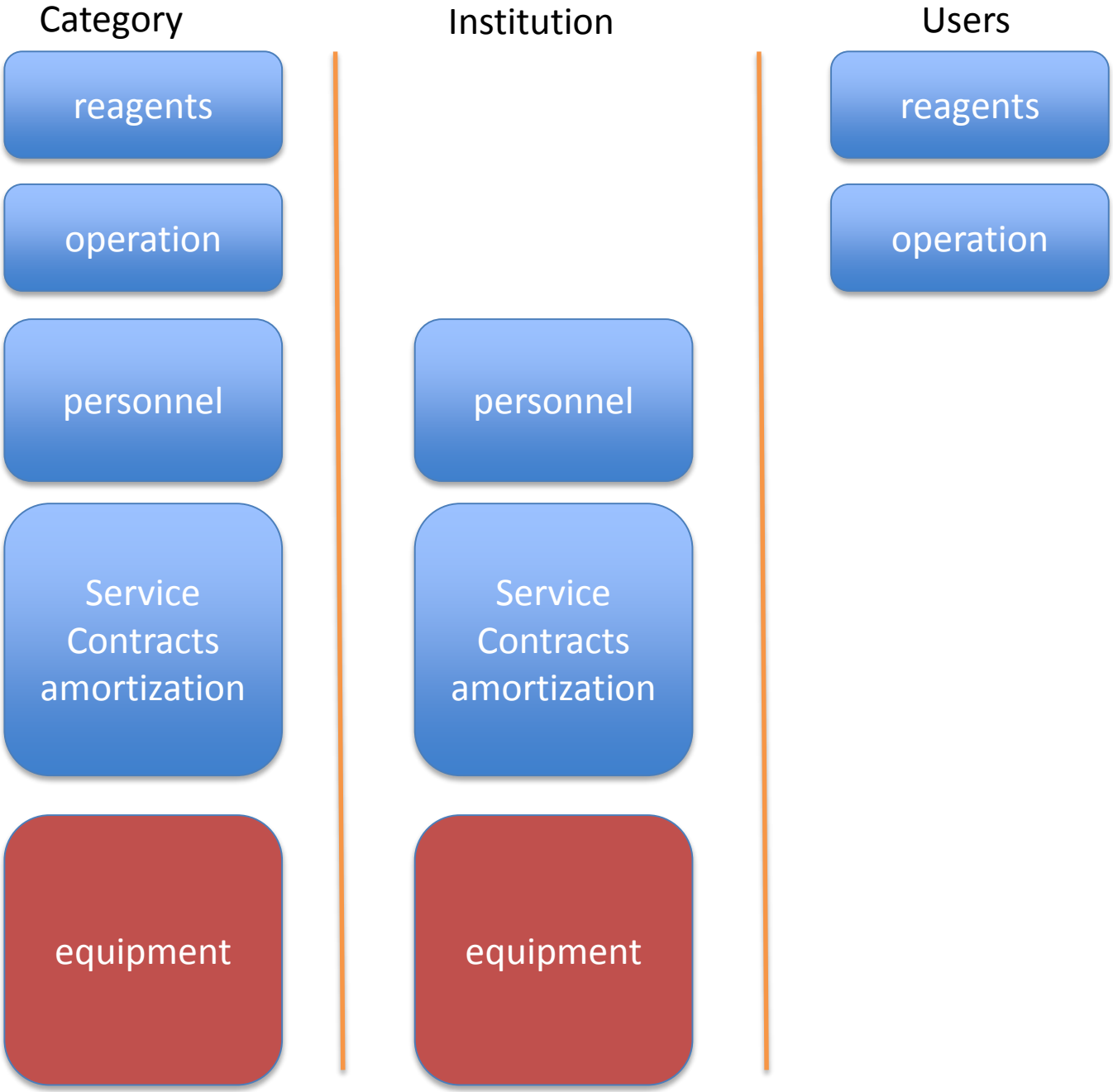
High Throughput Sequencing Facility Mission

- Be one of the best high-throughput genomic centers at public university.
- Support a cutting-edge research environment by providing broad, affordable access to the latest genomic technology.
- Help researchers to design and interpret genomic experiments; promote innovation by developing new and novel techniques
- Integrate the facility with other technology centers at the University of North Carolina; form partnerships within and beyond the UNC system

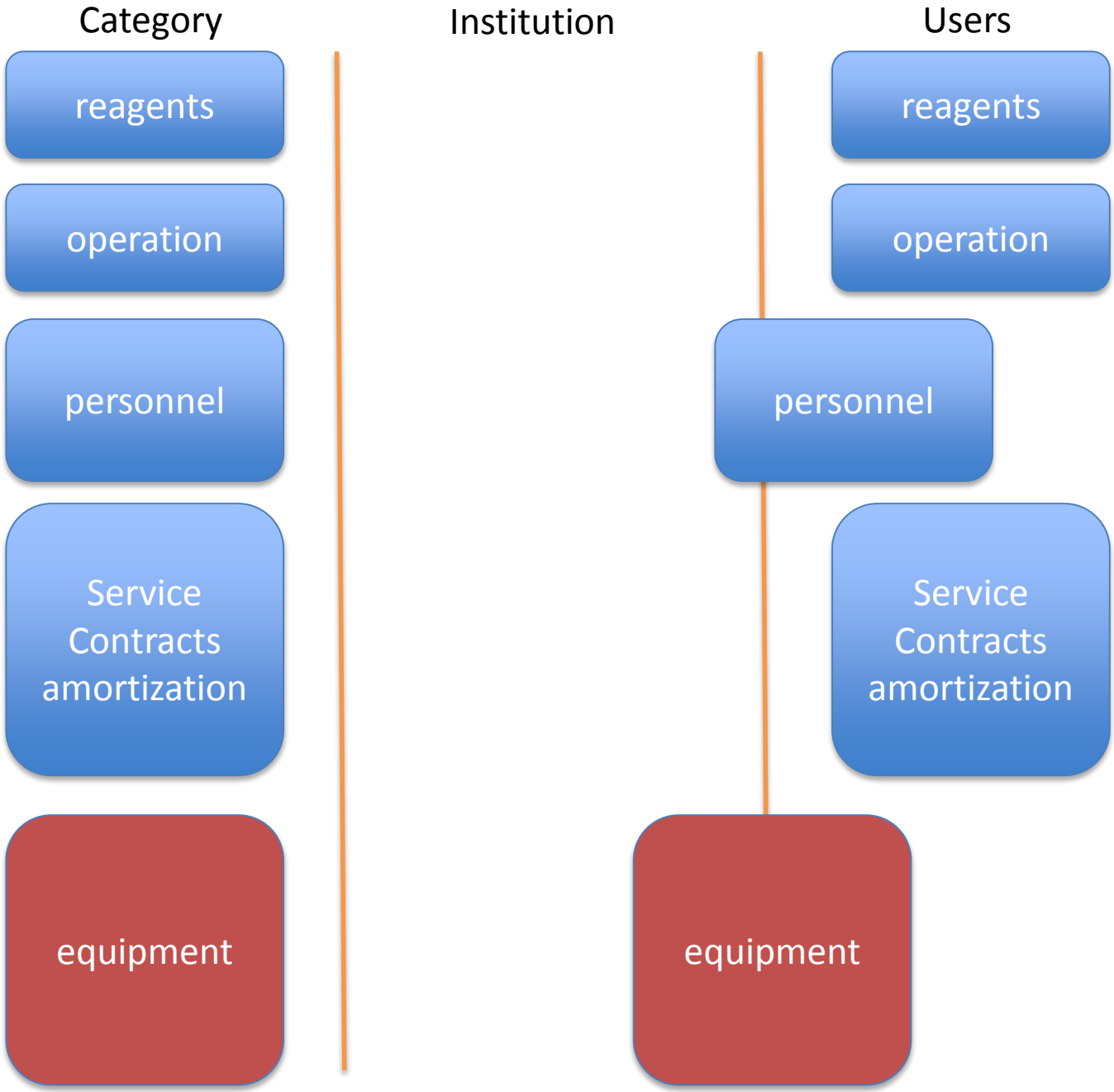
Initial start up for Facility development (2008)



Structure of Facility Funding in 2008-2010



Structure of Facility Funding in 2013



Increased number of instruments require more investment in the infrastructure

2008

Electrical Power
Efficient Air-condition system
Very Fast Internet Connection

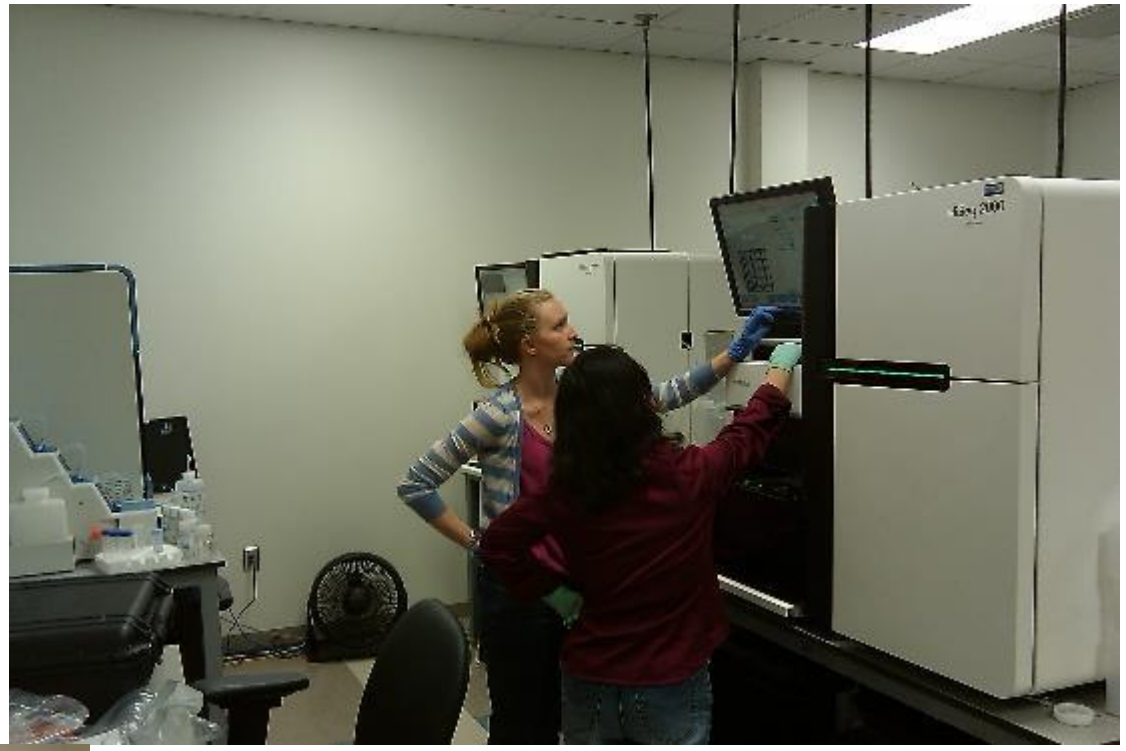
2013

Upgrades of equipment are important.

Do not be too attached to old equipment

UNC HTSF 2014

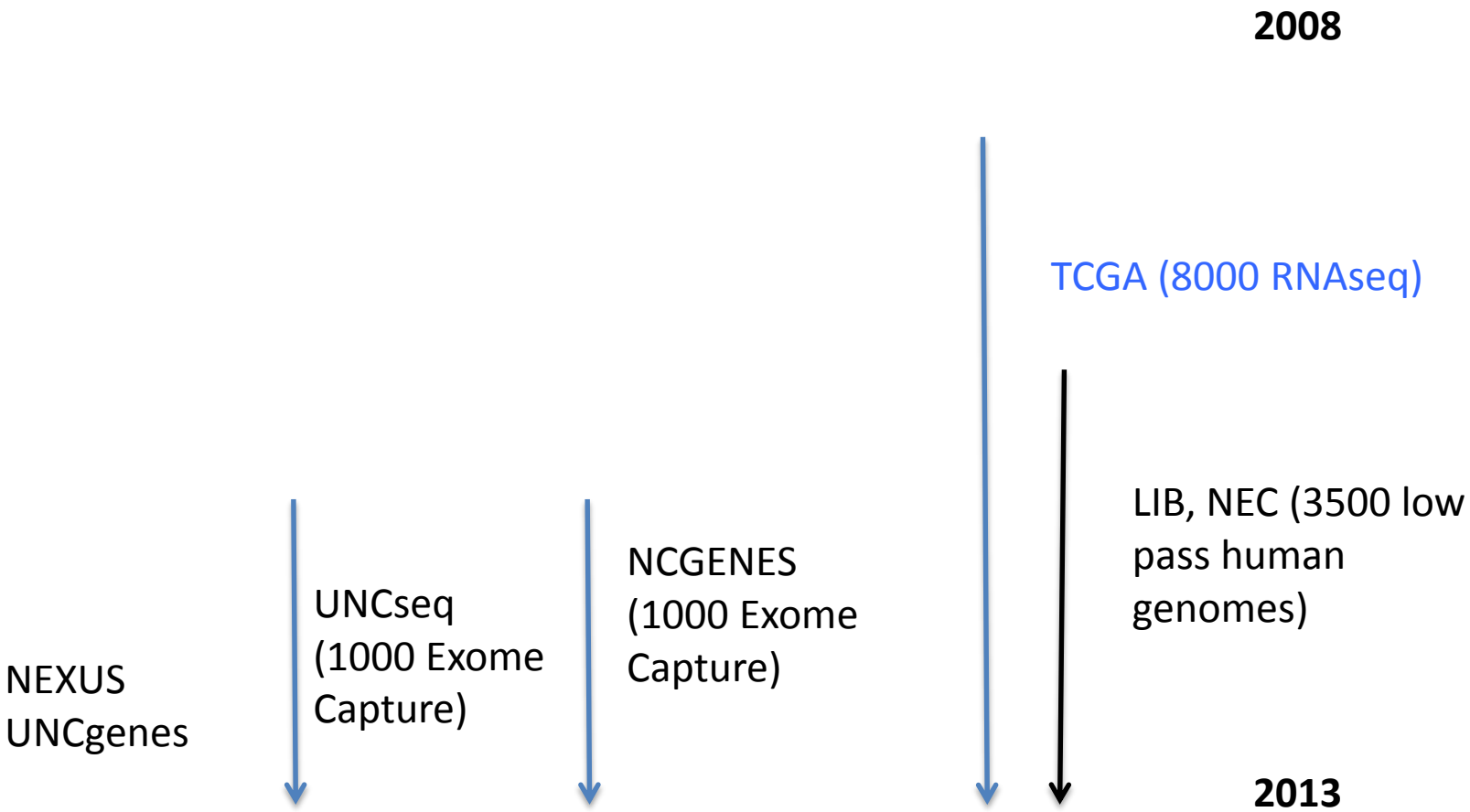
- 6 HiSeq 2500
- 4 HiSeq 2000
- 1 PacBio
- 1 PGM Ion Torrent
- 2 Ion Proton
- 3 MiSeq



Also on campus:

454 (Microbiome)
454 jr. (Viral genomics)
2 MiSeq
2 NextSeq500

Number of sequencers is always associated with increasing number of large studies



Total around 6000 samples in 2013

Projects

TCGA (The Cancer Genome Atlas project) - Gene Expression Patterns in Human Tumors Identified Using Transcript Sequencing
The goal of this project is the study of genome-wide transcript regulation with chromatin organization to provide a critical portrait of the cancer genome that can be integrated with other data, including mutations and copy number events.

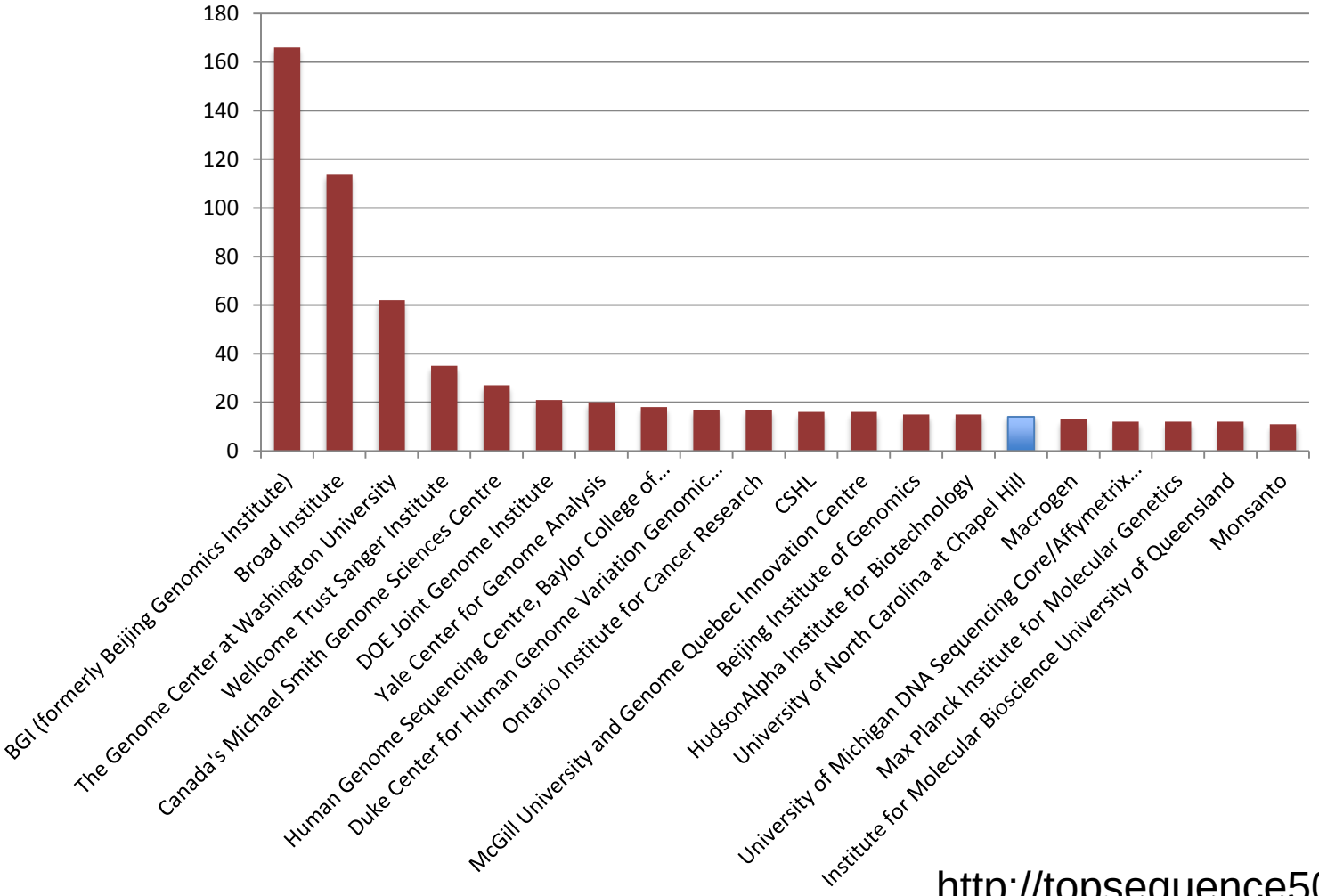
LIB, NEC - Deep Sequencing Studies for Cannabis and Stimulant Dependence
The goal of this project is to identify sequence variants that affect cannabis and stimulant dependence.

NC GENES: North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing.
The goal of this project is to establish a set of best practices to guide future implementation of robust genomic technologies

NC NEXUS, North Carolina Newborn Exome Sequencing for Universal Screening
The goal of this project is to identify, confront and overcome challenges to implement deep sequencing technology to enhance current newborn screening.

40% sequencing capacity for medium and small projects

World Rank 2011 -14th position
World Rank 2013 -15th position
Second largest NGS facility on East Coast



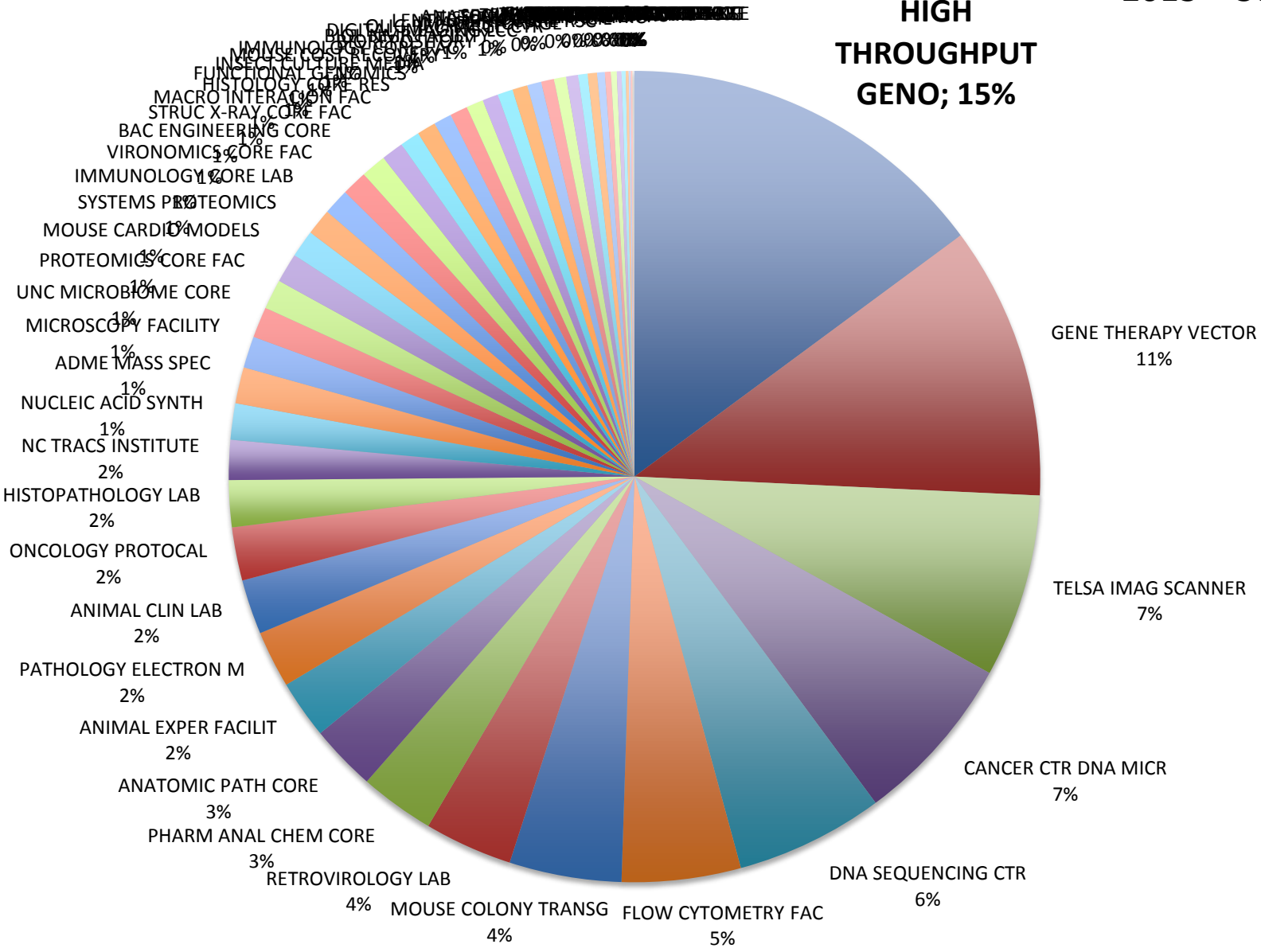
http://topsequence500.org/

Cores by Recharge Revenue

2011

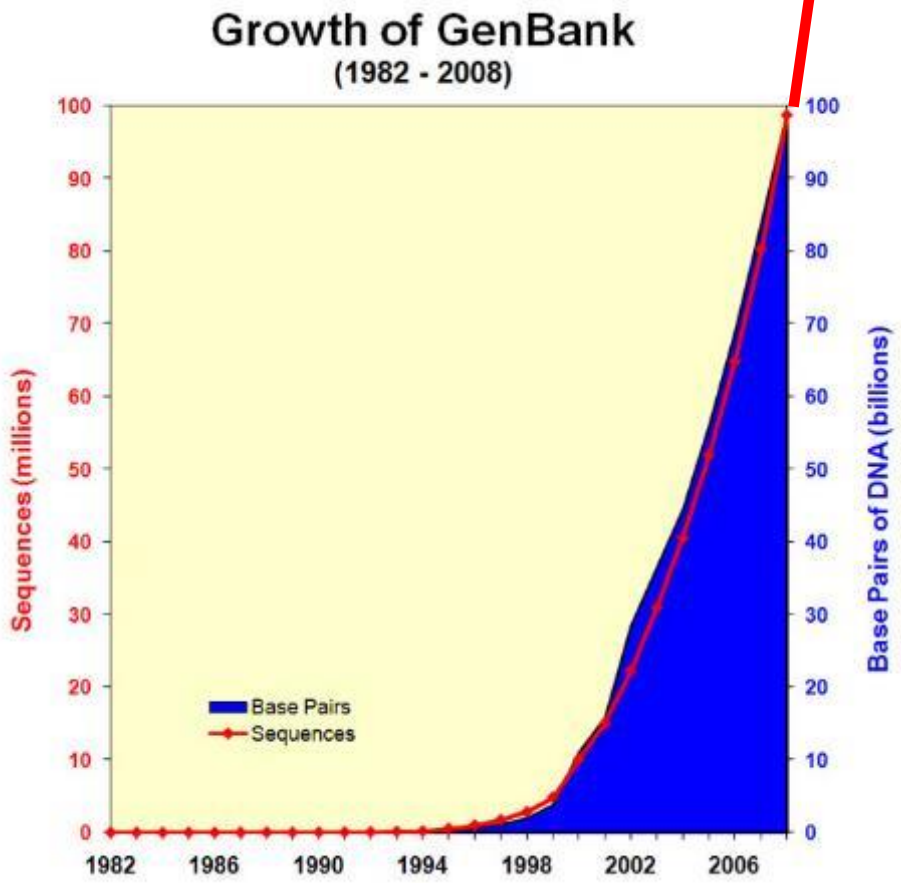
HIGH THROUGHPUT GENO; 15%

2013 – 30%



Capacity of HTSF

3,500,000,000,000 bp/week
UNC HTSF



Bioinformatics pipeline

Data Server

Data Analysis Cluster

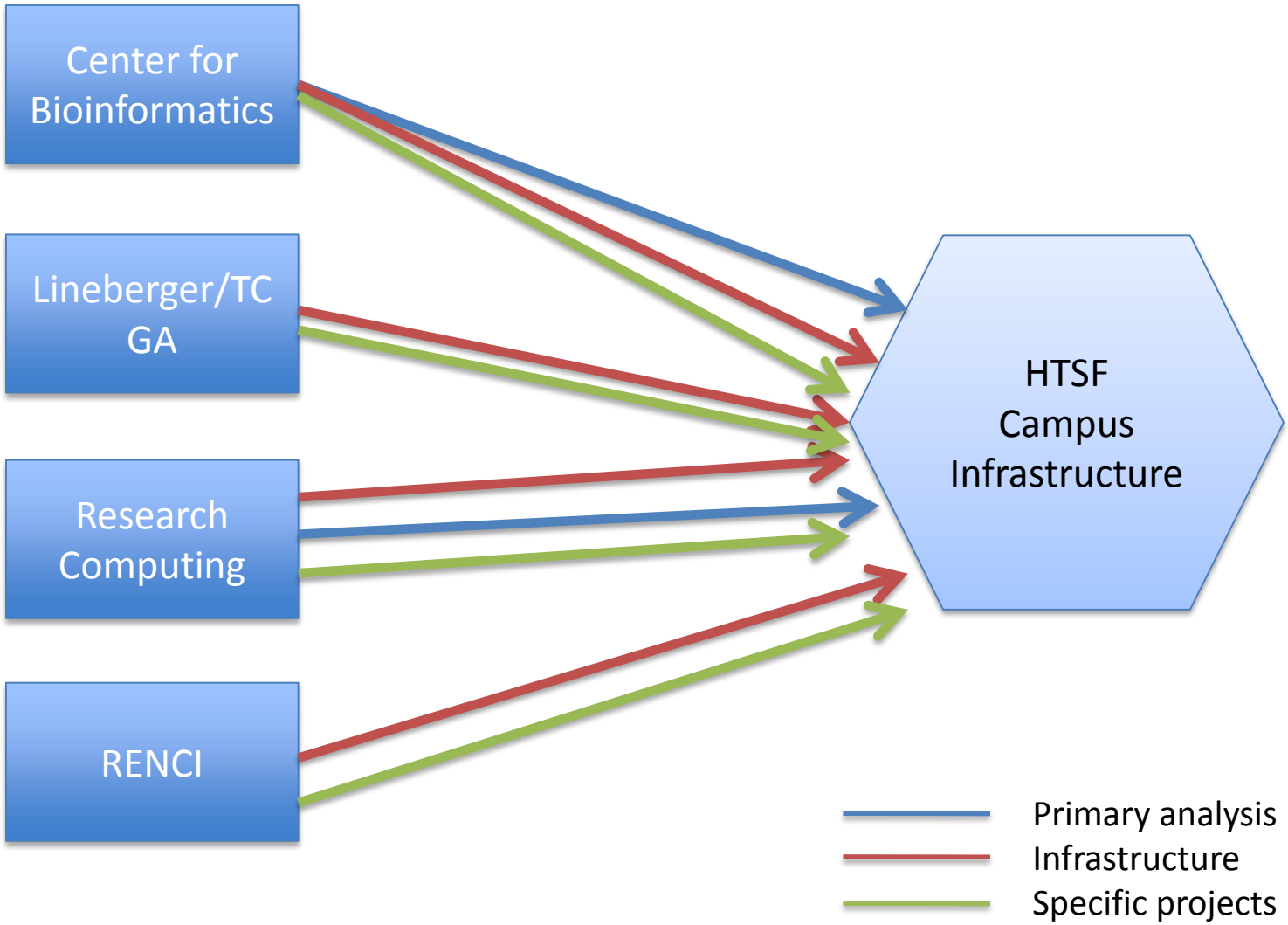
Data Analysis Cluster

Data Analysis Cluster

Data Storage 6-12 months

Tapes 3 years

Unified bioinformatics



Error rate

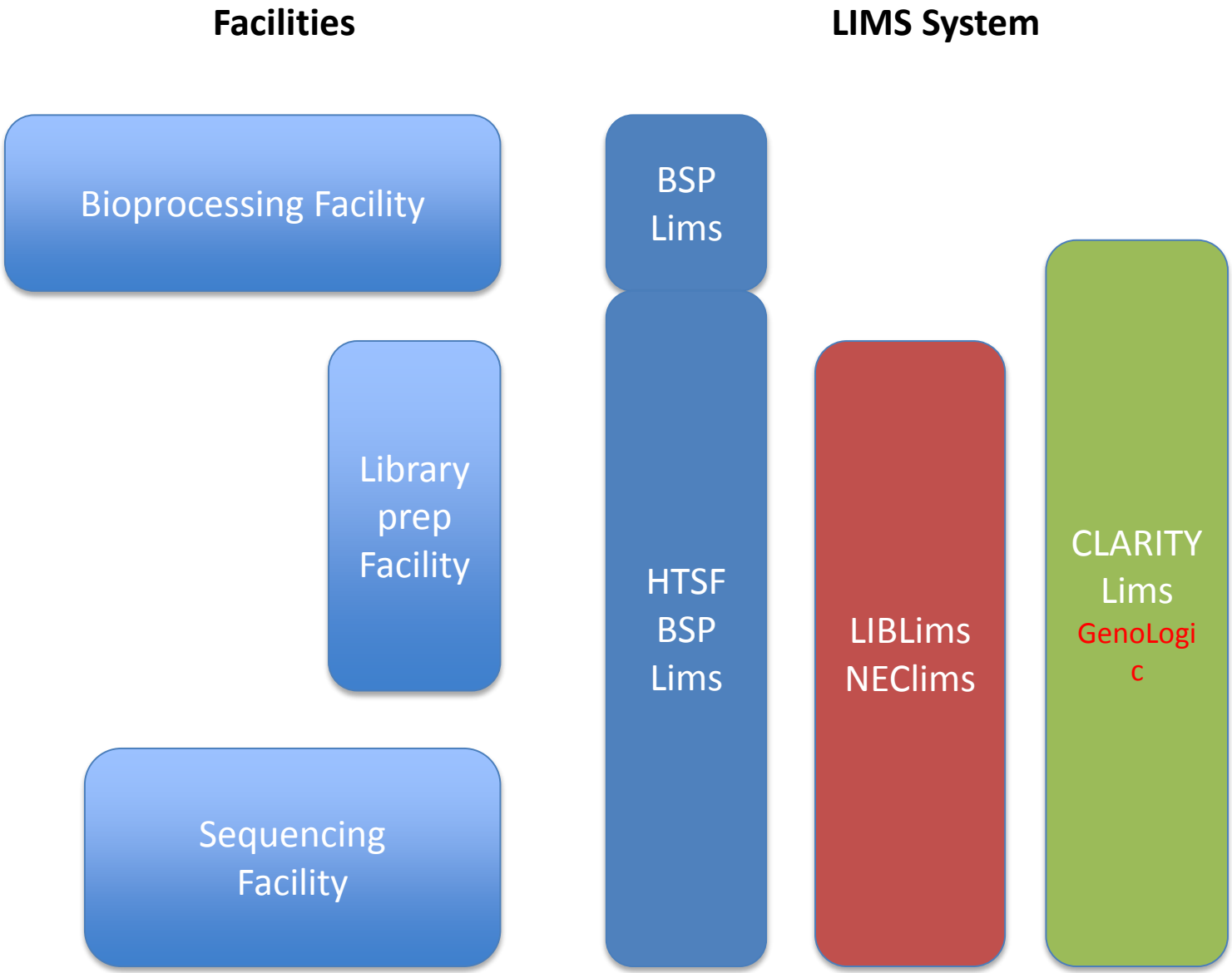
The more samples are processed, the chance for sample swap is greater.

Large projects are using SNP (microarray, sanger sequencing, Sequenom) for sample identification check.

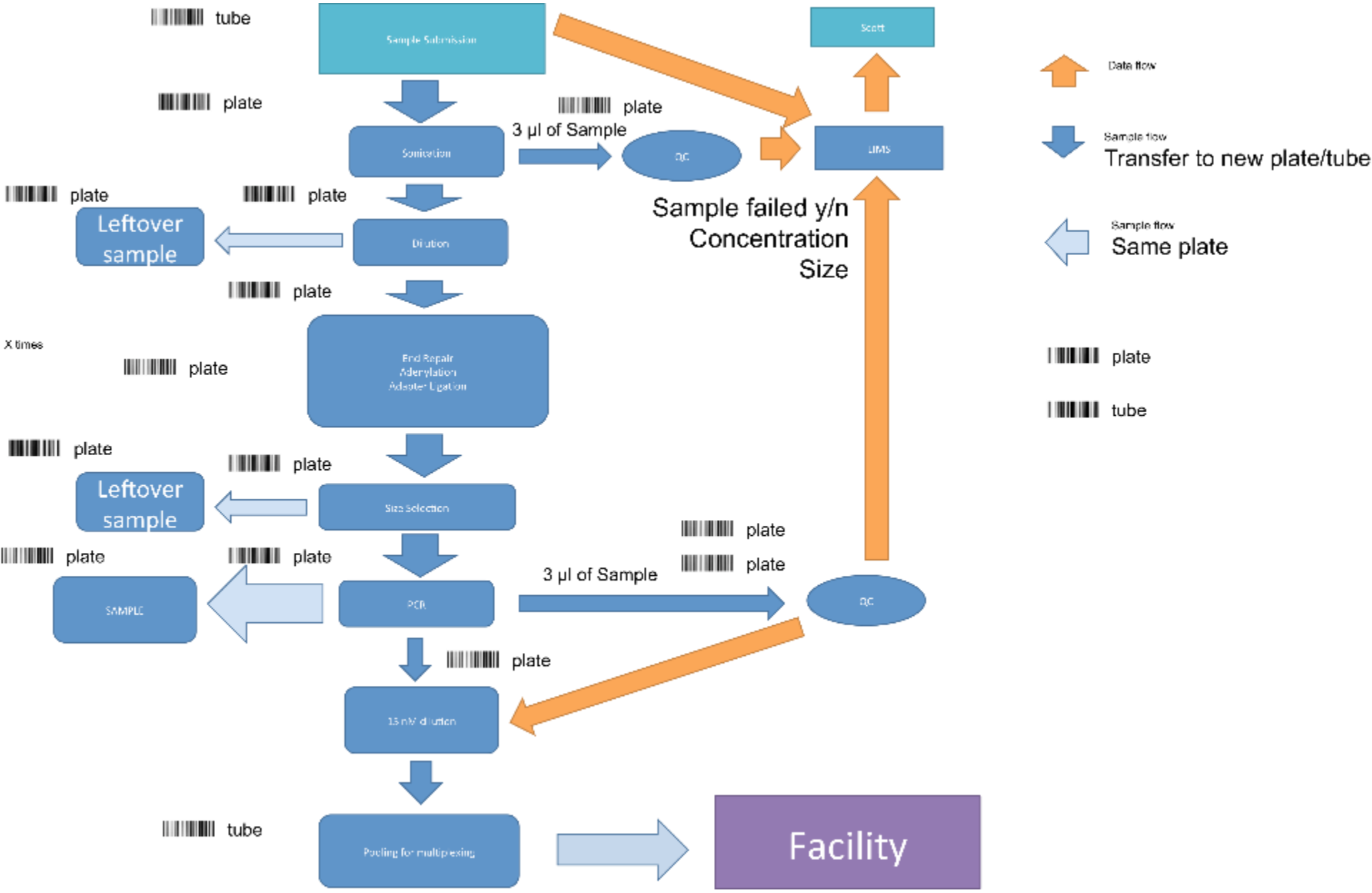
1. It is becoming necessary to use Laboratory Information Management System (LIMS) to control flow of the samples and reagents in the facility.
   Organized storage for incoming and processed samples.

2. Automation is essential for maintenance growth and reproducibility without increased number of sample swap cases.

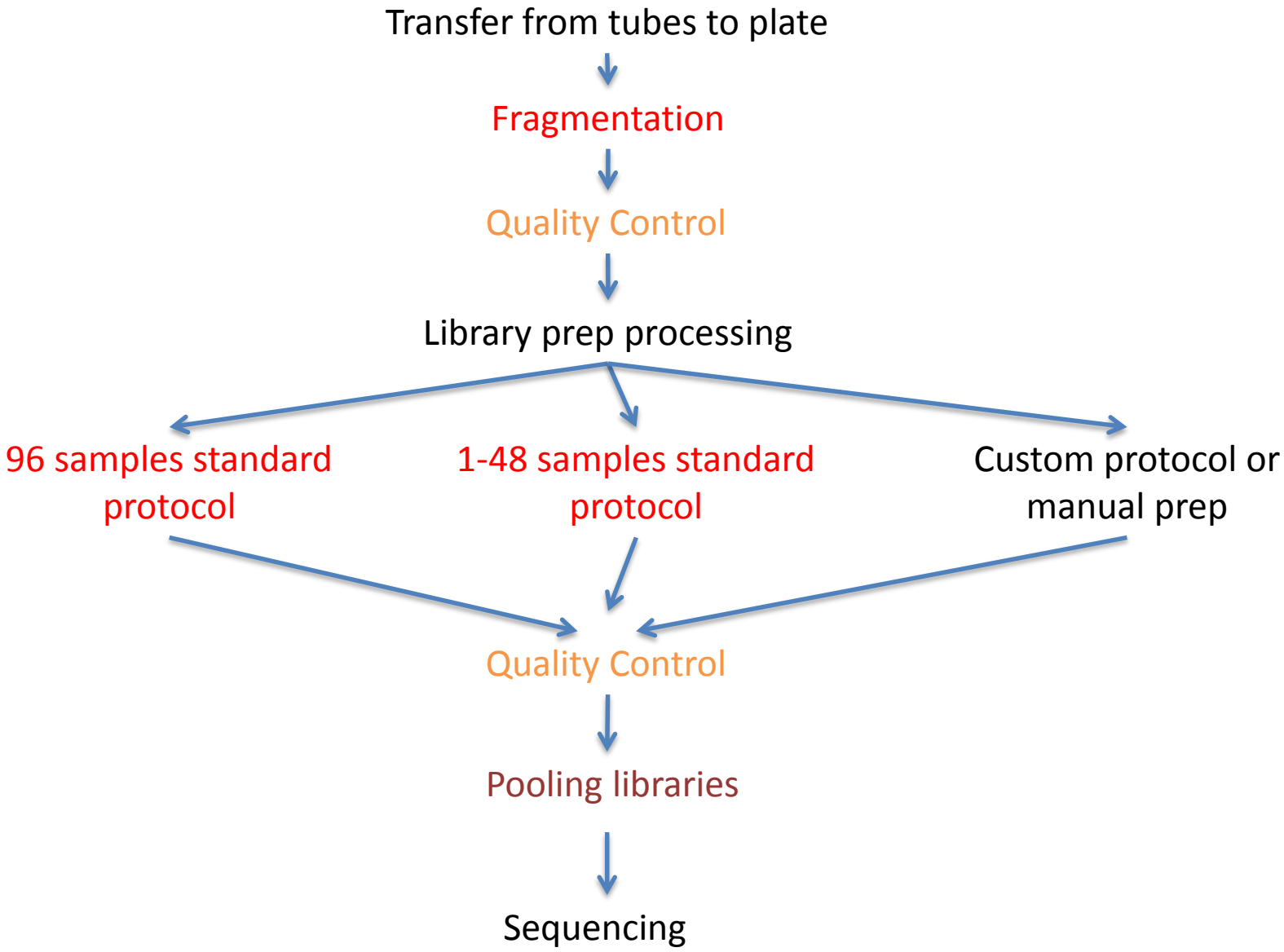
1. LIMS


Facilities
LIMS System
Bioprocessing Facility
Library prep Facility
Sequencing Facility
BSP Lims
HTSF BSP Lims
LIBLims NEClims
CLARITY Lims GenoLogic

BSP LIMS Library Preparation site



2. Automation



Automation in HTSF

Sonication

Medium and Large Scale library prep and DNA capture system

Caliper – Sciclone system
96 tip pipetting head

Agilent Capture protocols
KAPA DNA library protocol
Illumina DNA/RNA protocols

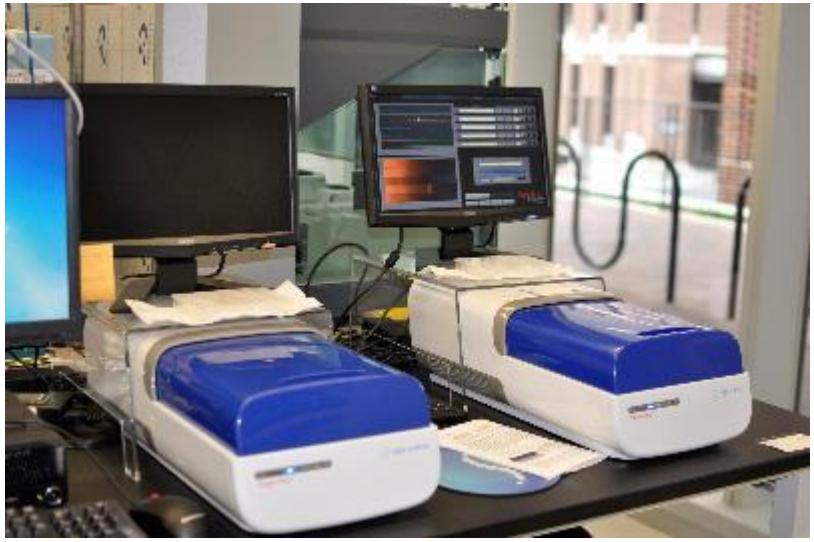
Automation in HTSF

Tecan – Freedom Evo system – 8 tips
- 2x48 (96) samples per week – DNA library prep
- Automated sample normalization steps
- PCR and qPCR preparation
- Integrated fluorescent plate reader

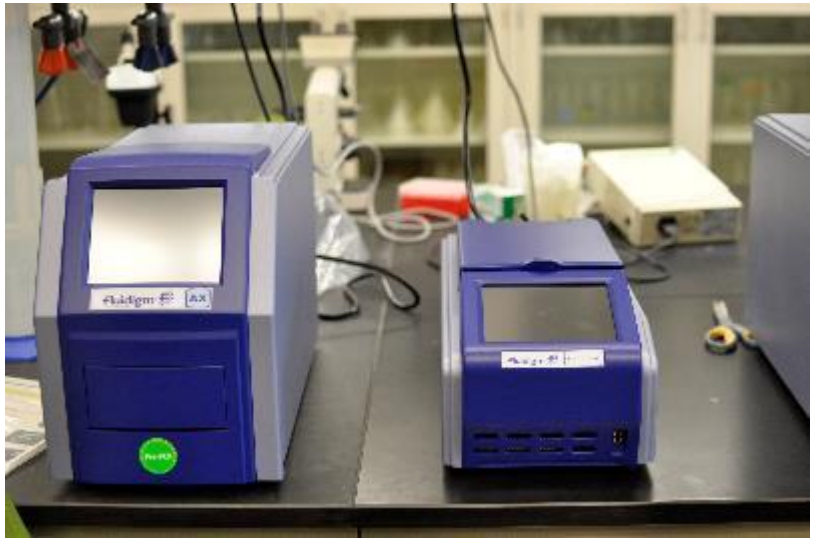
Infinite F200
- Can be adapted to small and medium scale protocols for Illumina and Ion Torrent
- We have all necessary components for DNA/RNA extraction using Qiagen kits

Tecan – Freedom Evo system – 8 tips and 96 tips
- For Illumina and Ion torrent protocols
- Experimental platform for new protocols

Automation in HTSF



Sage Science
2 x Pippin – Automated size selection system
1 x Blue Pippin – large fragment selection for PacBio protocol



Fluidigm
Access Array System – for large scale amplicon generation

Automation necessary for reproducibility

Automated Illumina Protocols on the Tecan NGS Workstation

Validated Protocols:

- Nextera DNA Sample Preparation
- Nextera Rapid Capture (Exome, Extended Exome, Custom)

During Validation:

- TruSeq Stranded mRNA Preparation
- TruSeq Stranded Total RNA Preparation
- TruSeq Nano DNA Sample Preparation
- Nextera XT DNA Sample Preparation

During Script Development:

- TruSeq DNA PCR-Free Sample Preparation
- TruSeq SmallRNA Sample Preparation
- TruSeq ChIP Sample Preparation
- TruSeq Custom Amplicon Library Preparation

Additional Protocols generated on Tecan NGS Workstation

- KAPA Hyper Kit
- KAPA NGS quantitation kit
- Amliseq – Life Technologies

Tecan – Illumina - UNC

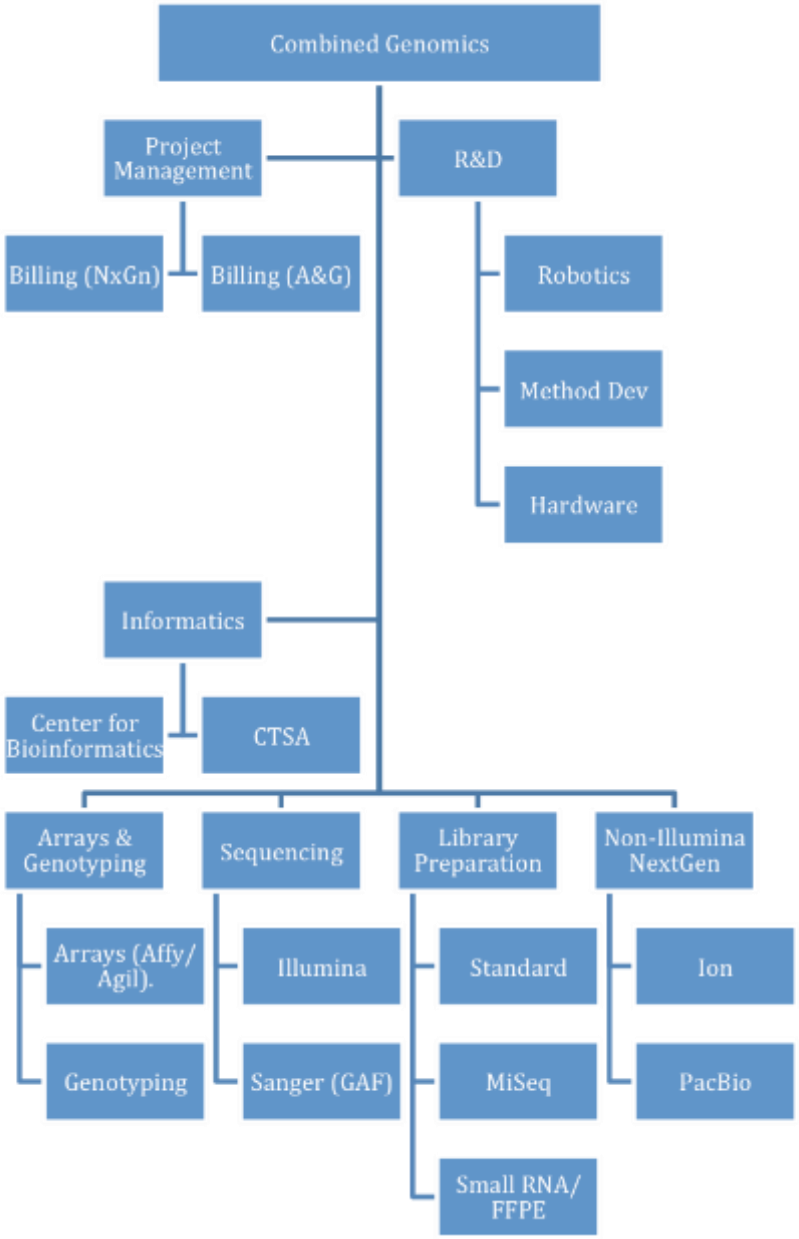
Research and Development are important for accessing the newest technologies

- Testing of the new chemistry – sequencing and library prep (Illumina)
- Testing of the new hardware
- Research on chemistry for Molecular Tagging
- Custom library preps according to user request
- Automation protocols (Tecan, Illumina, KAPA, Rubicon Genomics)

People

- Trained staff is critical for high quality of the service.
- We are always keeping employment on minimum level (but on competitive salaries).
- Training and attending conferences are integral part of staff education.

Reorganization of Genomics Cores at UNC - 2014



Conclusions

- Initial startup should be designed well.
- Development of the facility should be correlated with the presence of stable, large projects.
- Infrastructure (space, electrical power, air-conditioning and such) needs to satisfy the current equipment requirements and provide the room for future additions.
- Flexibility in protocols and collaboration with users is essential.
- LIMS and Automation – provide the control over errors.
- People – are the key to success.

Acknowledgements

Mieczkowski Lab + HTSF
Ewa Malc
Donghui Tan
Liz Sheffield
Maryam Clausen
Alicia Brandt
Nick Schuch
Uma Veluvolu
Scot Waring
Tara Skelly

Hemant Kelkar
Tristan De Buysscher

Corbin Jones
Christopher Baker