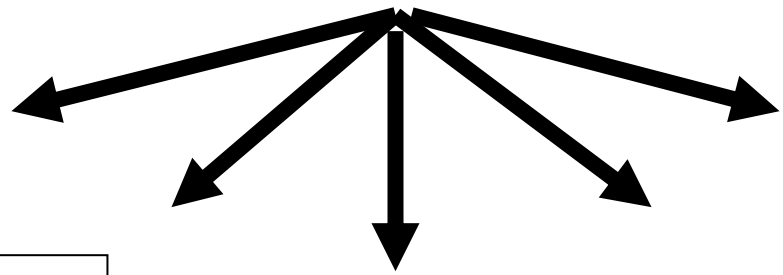
Shared Materials Repositories at the **Harvard Institute of Proteomics**

DF/HCC DNA Resource Core

Protein Structure Initiative Materials Repository (PSI-MR)

Protein Function

GLOBAL QUESTIONS



PROTEIN INTERACTIONS

STRUCTURE

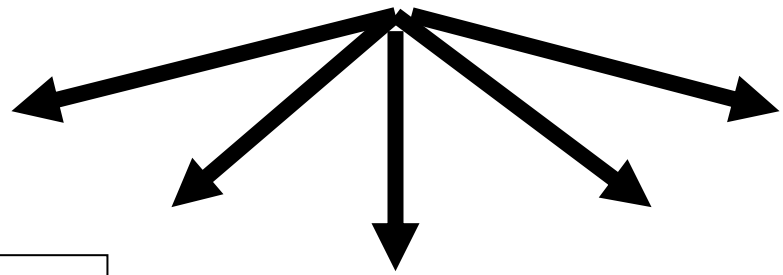
PROTEIN MODIFICATIONS

SUB-CELLULAR LOCALIZATION

ENZYMATIC ACTIVITY

Protein Function

GLOBAL QUESTIONS



PROTEIN INTERACTIONS

STRUCTURE

PROTEIN MODIFICATIONS

ENZYMATIC ACTIVITY

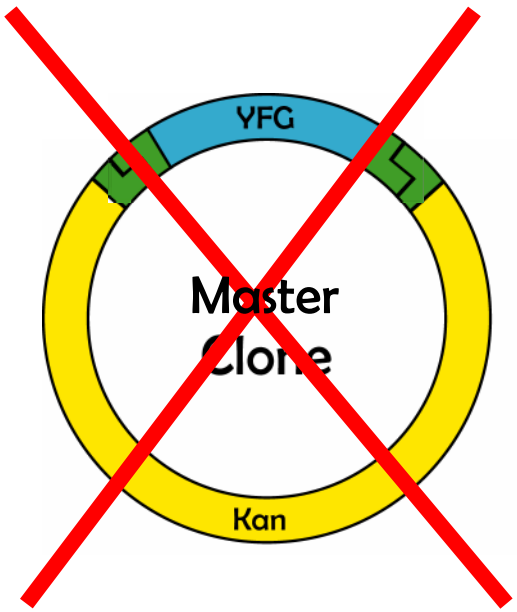
SUB-CELLULAR LOCALIZATION

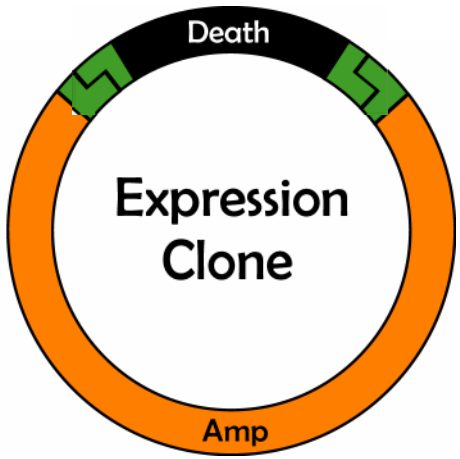
- **Express proteins in Different Cellular Contexts**
- **Express the Full-Length Protein**
- **Express with Appropriate Tags**

High-Throughput Protein Studies

Problems	*Solutions*
Express in different cellular contexts	*Ability to put gene in different vectors*
Express with appropriate tags	
Express full-length proteins	*Verification of gene insert sequences*

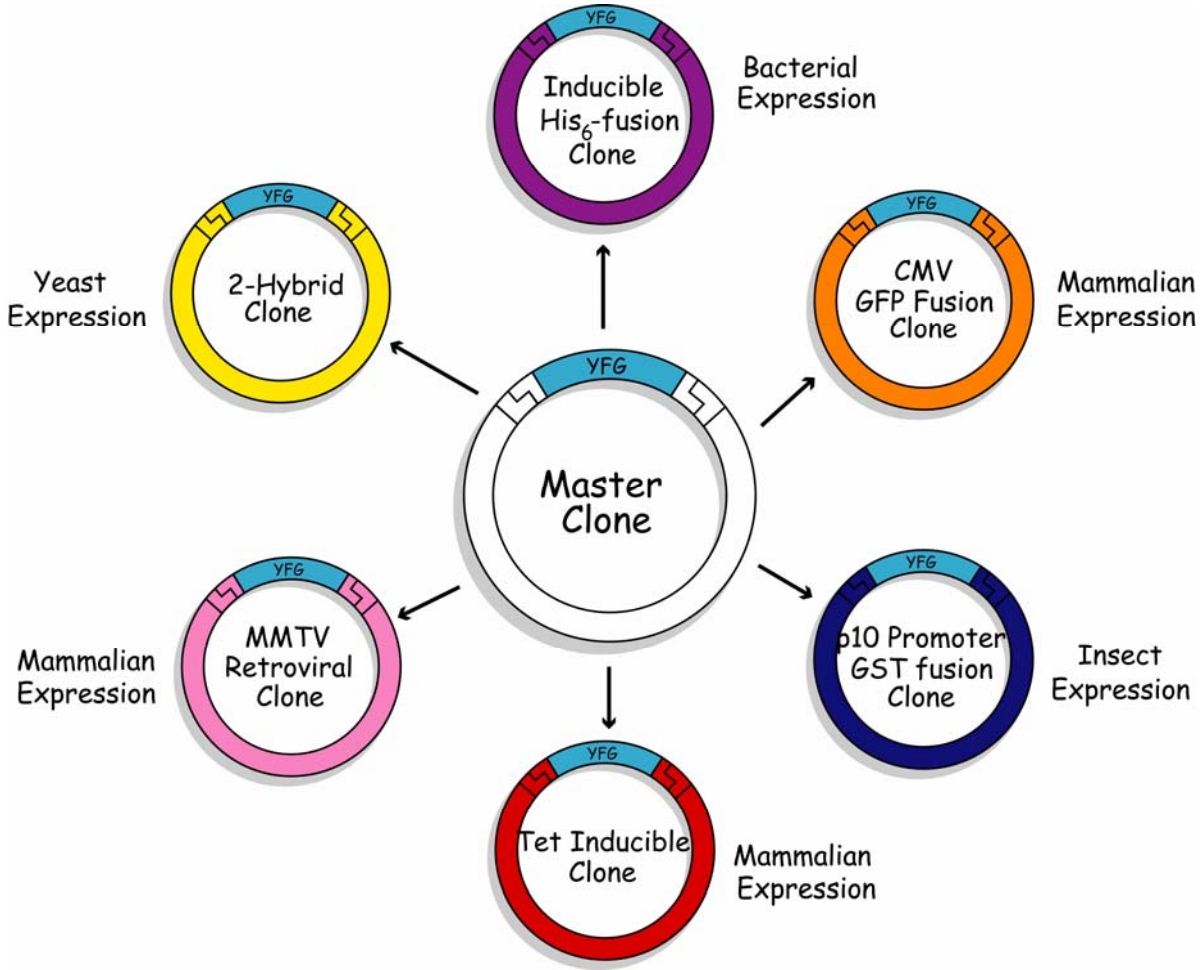
Moving Genes by Recombination

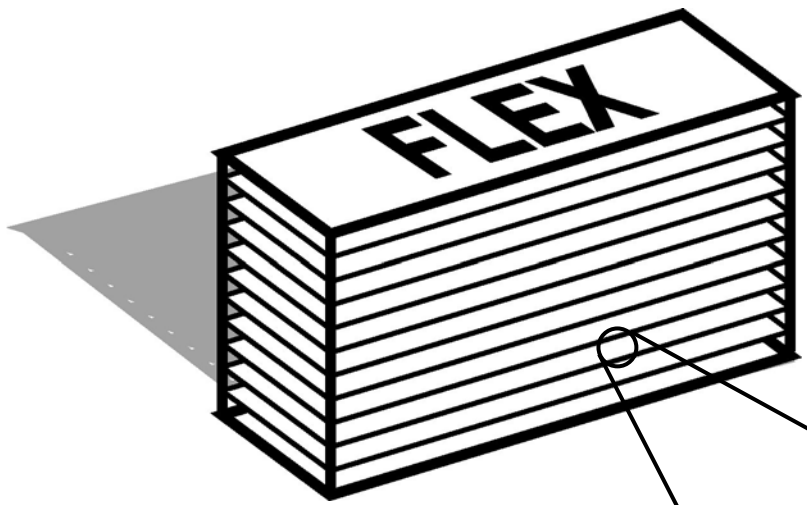




Only the Desired Product Survives

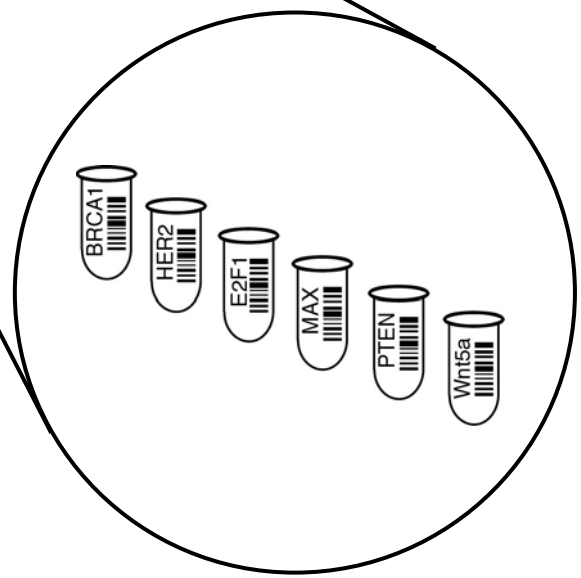
Testing Proteins Many Different Ways



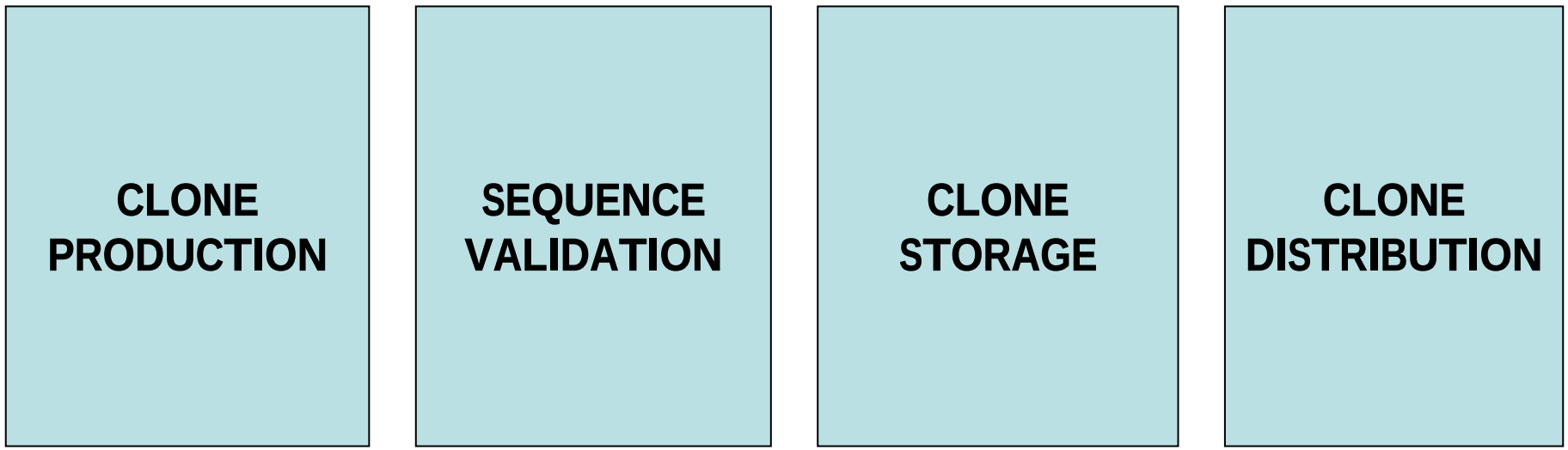


All genes
Broadly available
No Restrictions
Flexible format
Expression-ready
Sequence-verified
Affordable

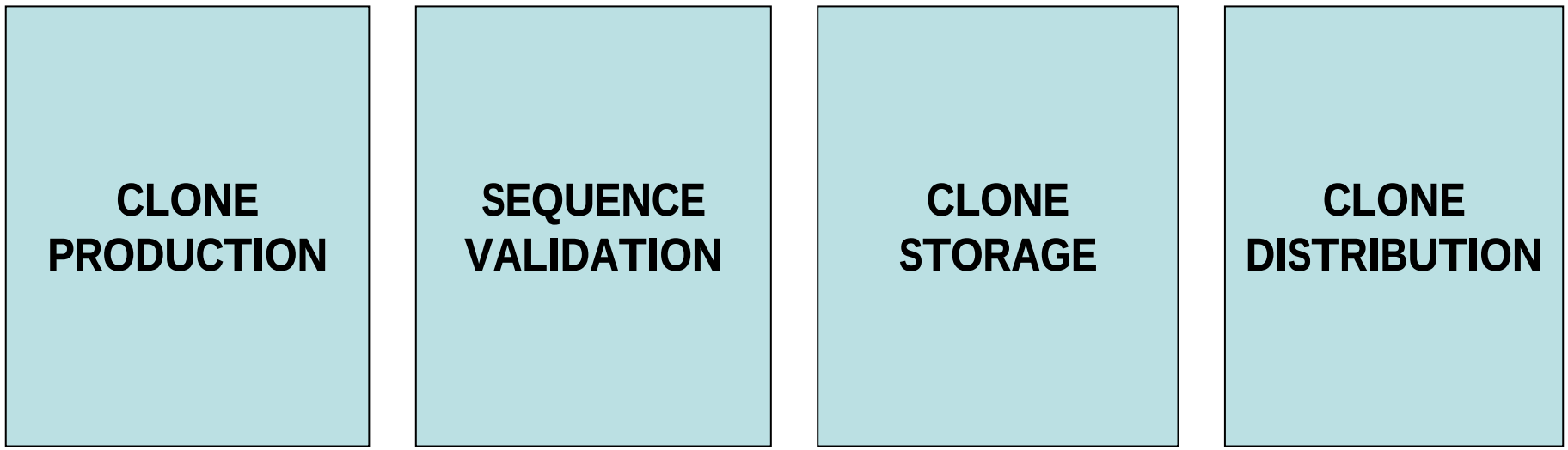
Each tube
One full-length coding region



Clone Collections—Issues



Clone Collections—Issues



- **Informatics**
- **Automation**

Clone Collections—Issues

CLONE PRODUCTION	SEQUENCE VALIDATION	CLONE STORAGE	CLONE DISTRIBUTION

<u>Any</u> Plasmid Clone

- **HIP FLEXGene Clones**
- **DF/HCC DNA Resource Core**
- **PSI Materials Repository**

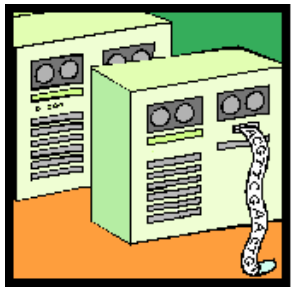
DF/HCC

PSI Protein Structure Initiative

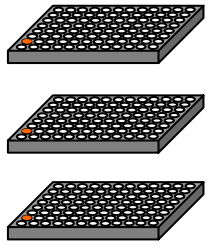
Clone Collections—Issues

Select genes

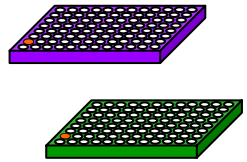
Design Primers



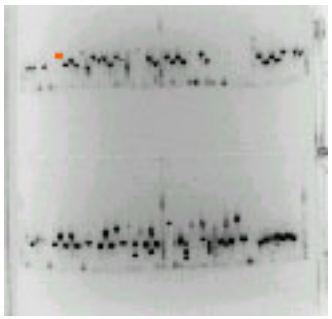
Receive Primers



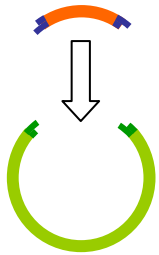
PCR



Gel Purify



Clone



Transform

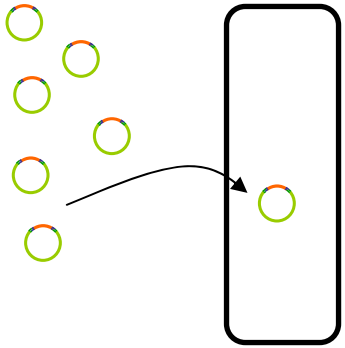
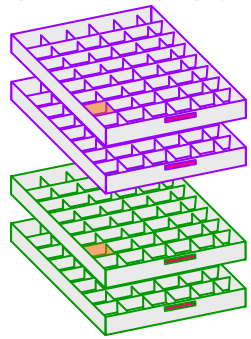
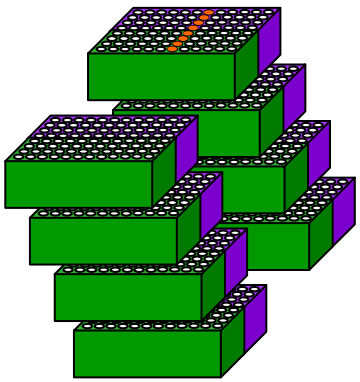


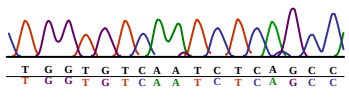
Plate for Selection



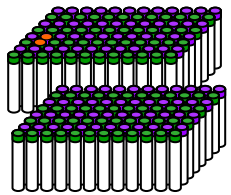
Pick/Culture



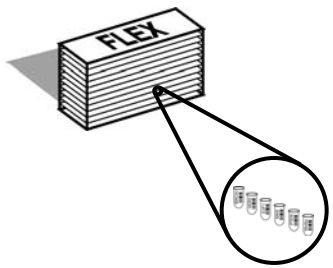
Sequence Verify



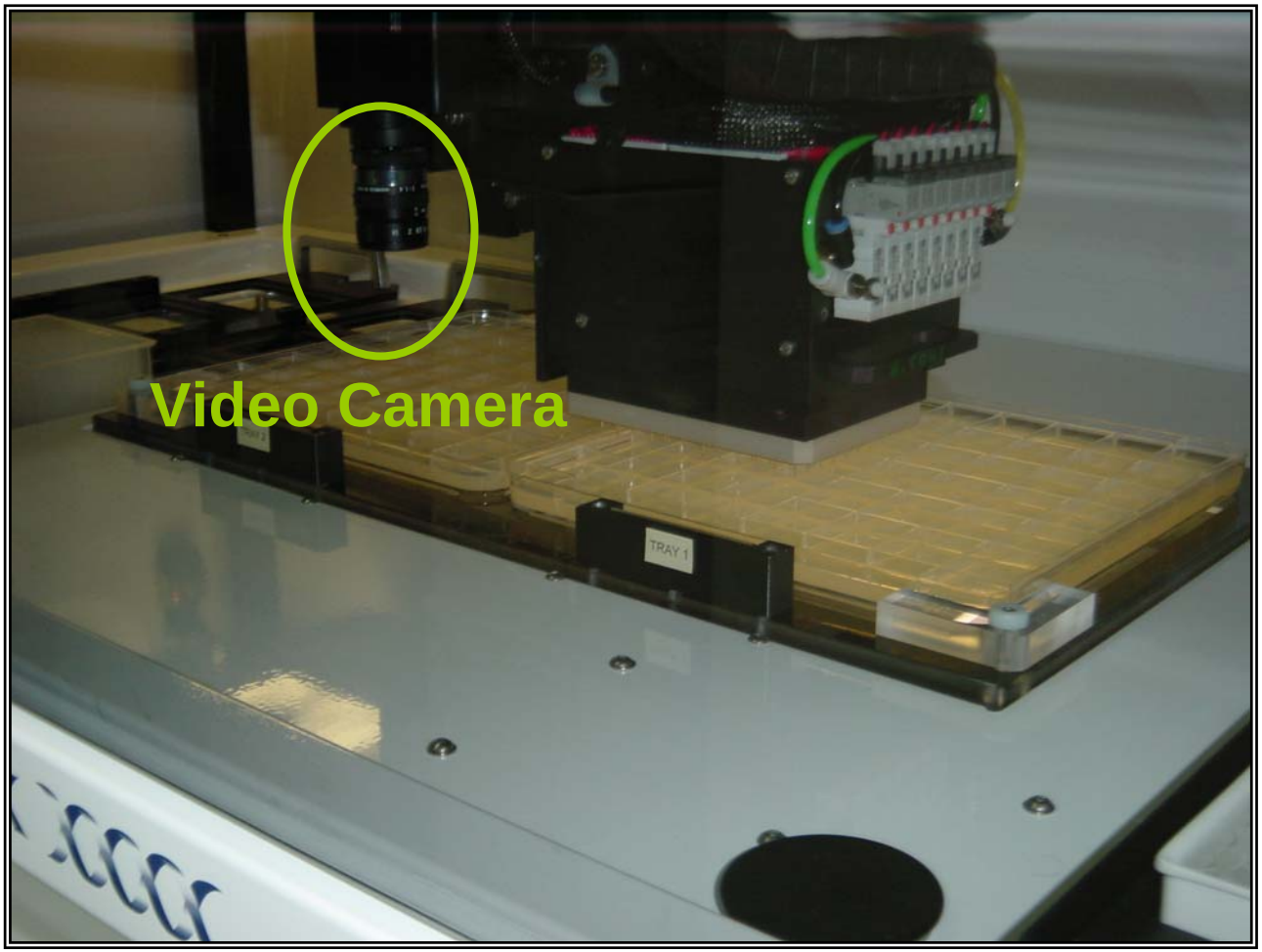
Hit Pick/Prep



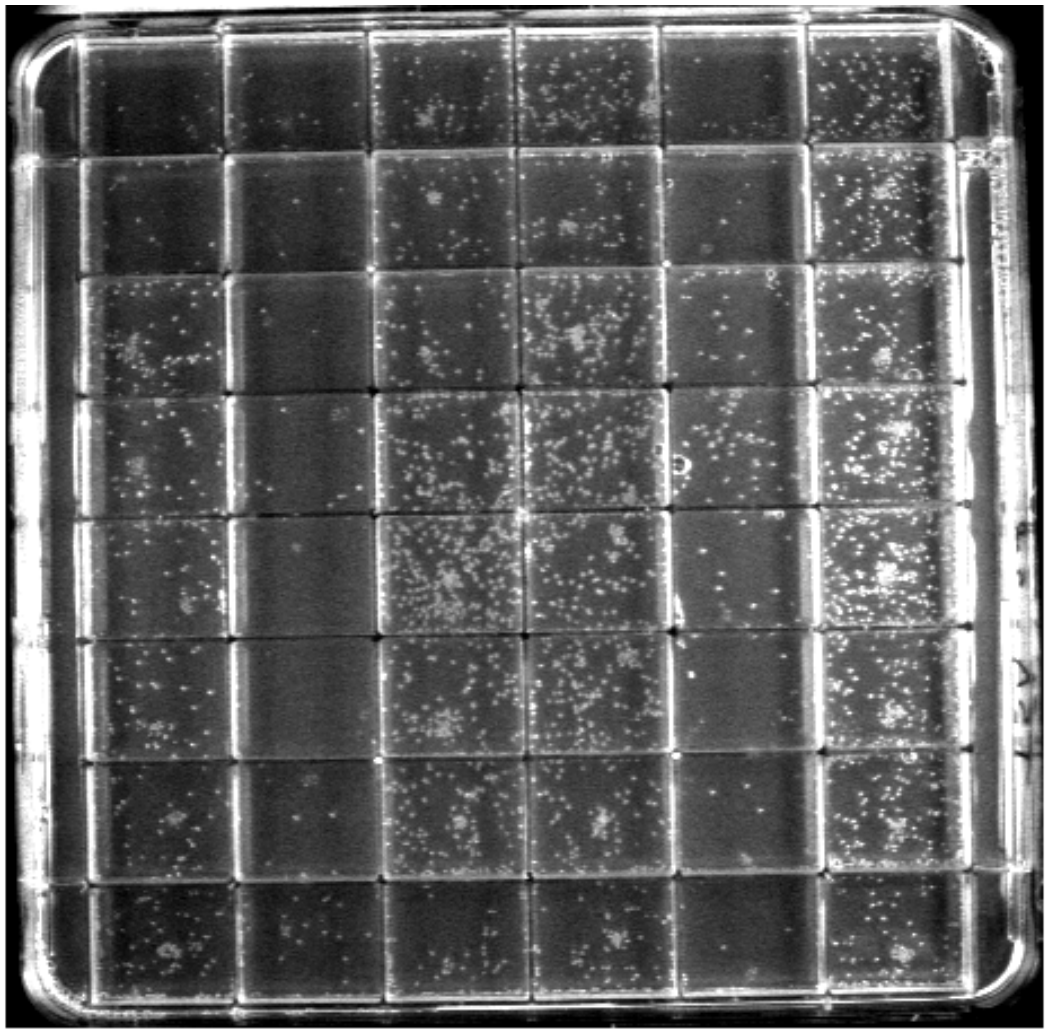
Add to FLEXGene



Automated Colony Picking

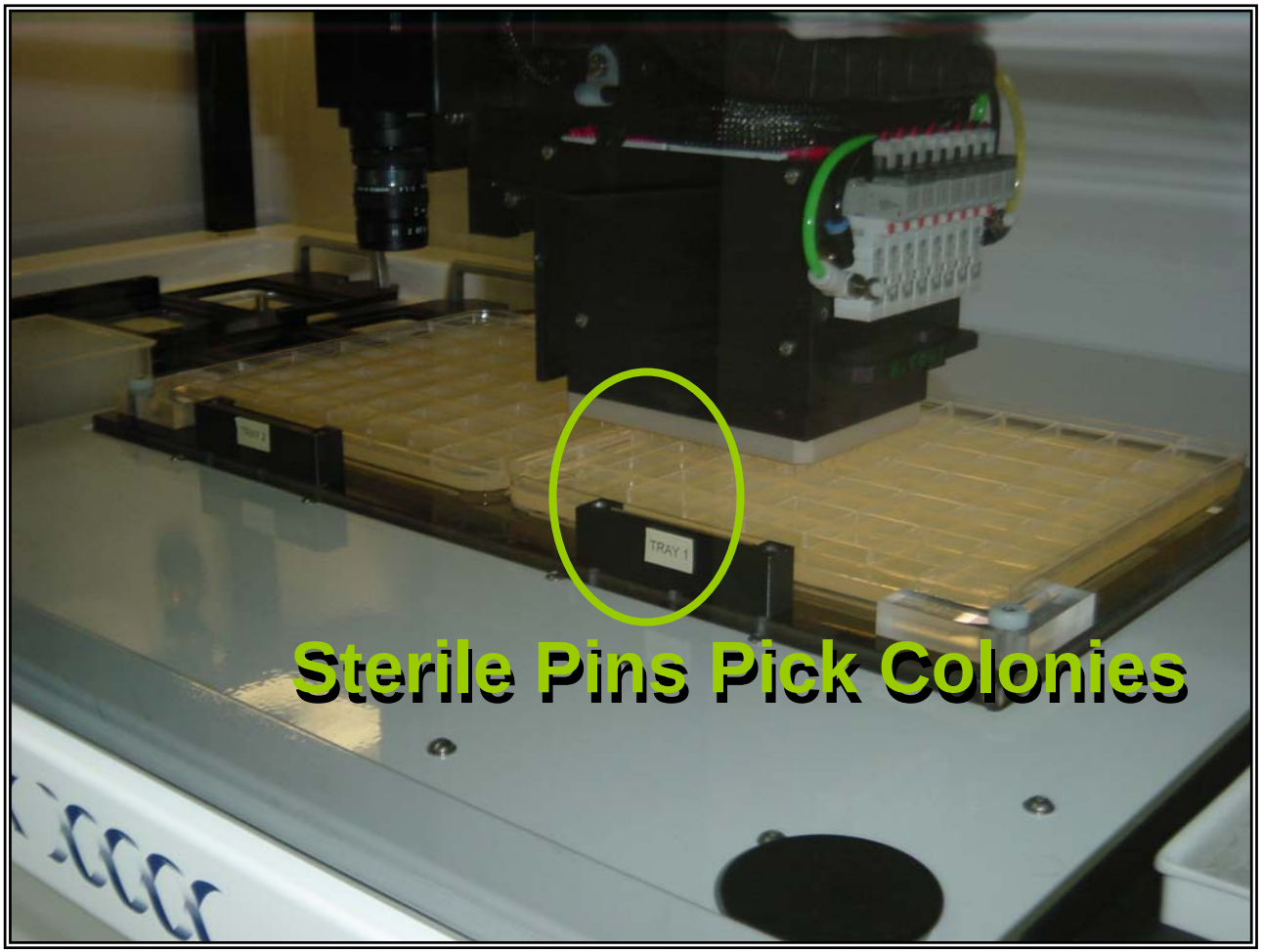


Plating Bacteria



New plate design allows 48 different clones to be plated on the same dish

Automated Colony Picking

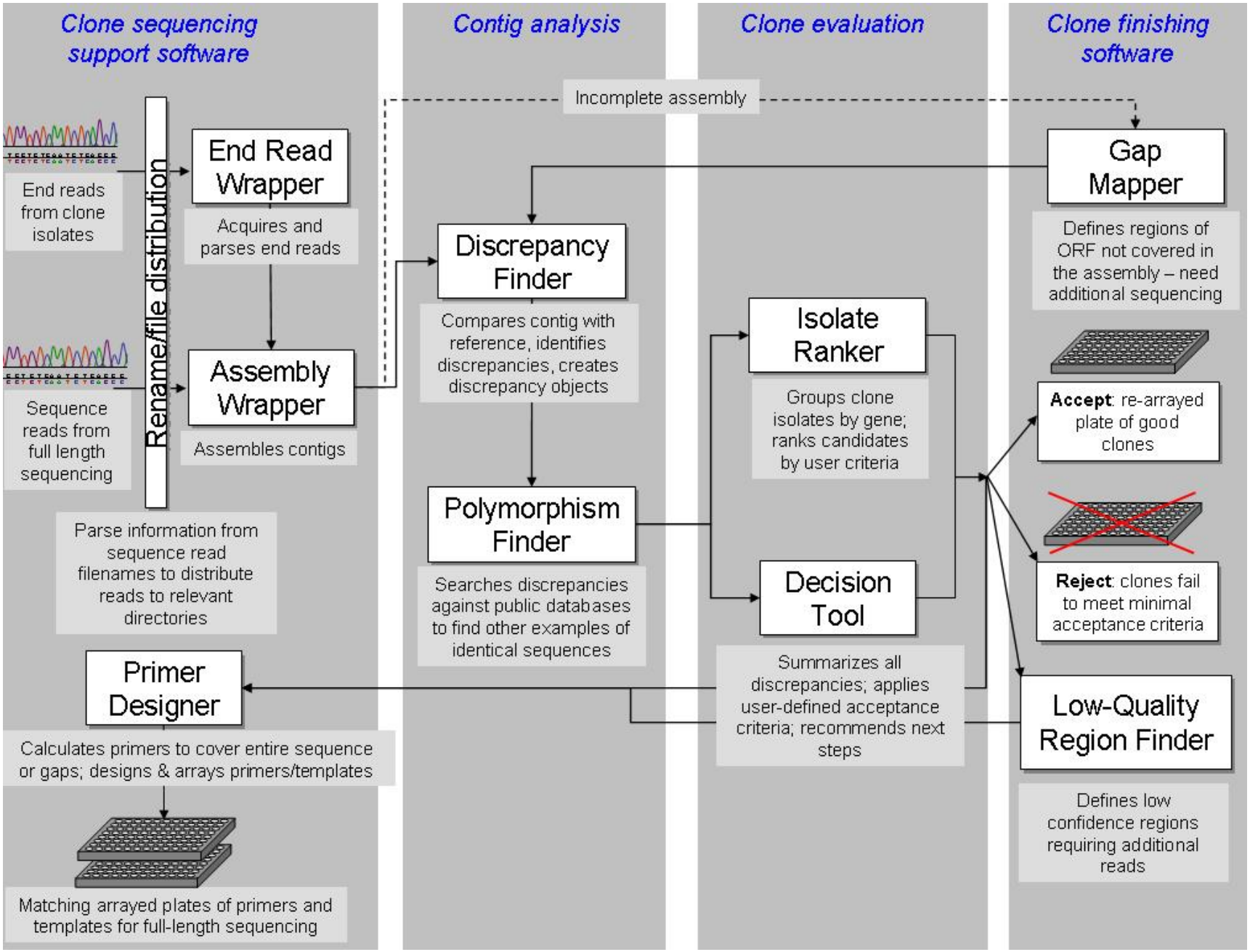


FLEXGene Clone Production

- Hundreds of clones per week
- Quality control measures
  - Single-colony selection
  - Phage-resistant host strains
  - Barcode label tracking
- Fully tracked in the FLEXGene LIMS
- Routine protocols for transfer to expression vectors

Clone Collections—Issues

- ACE software tool
- ABI 3730 DNA Sequencers


Clone sequencing support software
Contig analysis
Clone evaluation
Clone finishing software
Incomplete assembly
End reads from clone isolates
Rename/file distribution
End Read Wrapper
Acquires and parses end reads
Sequence reads from full length sequencing
Assembly Wrapper
Assembles contigs
Parse information from sequence read filenames to distribute reads to relevant directories
Discrepancy Finder
Compares contig with reference, identifies discrepancies, creates discrepancy objects
Polymorphism Finder
Searches discrepancies against public databases to find other examples of identical sequences
Isolate Ranker
Groups clone isolates by gene; ranks candidates by user criteria
Decision Tool
Summarizes all discrepancies; applies user-defined acceptance criteria; recommends next steps
Gap Mapper
Defines regions of ORF not covered in the assembly – need additional sequencing
Accept: re-arrayed plate of good clones
Reject: clones fail to meet minimal acceptance criteria
Low-Quality Region Finder
Defines low confidence regions requiring additional reads
Primer Designer
Calculates primers to cover entire sequence or gaps; designs & arrays primers/templates
Matching arrayed plates of primers and templates for full-length sequencing

FLEX gene ORF collections

Species	**Unique genes cloned**
S. cerevisiae *	5,000 genes
H. sapiens	8,000 genes
P. aeruginosa	5,570 genes
Y. pestis *	4,200 genes
F. tularensis *	2,023 genes
V. cholerae *	3,836 genes
B. anthracis	3,500 genes

* Complete or near-complete coverage of the ORFeomes

Clone Collections—Issues

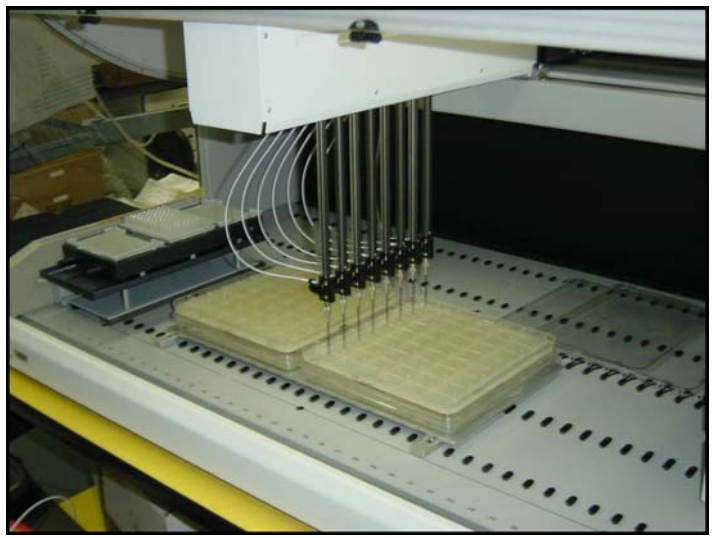
Clone Collections—Issues

- **FLEXGene clones—single colonies, phage-resistant hosts**
- **Other labs' clones—not necessarily same level of QC**

Storage Sample Production

Sample Storage Production

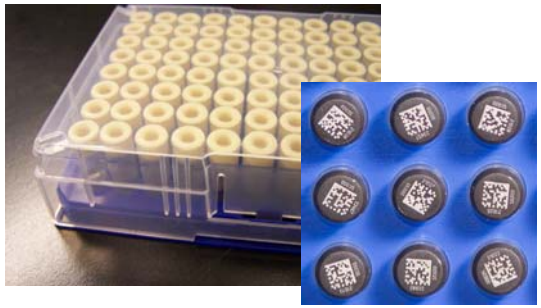
Samples arrive as DNA in solution (barcode-labeled plates)



Transformation & Robotic Colony Pick (barcode-labeled 48-sector dishes)

Liquid Culture (barcode-labeled deep-well blocks)

Sample Storage Production



Working sample
(barcode-labeled tubes)

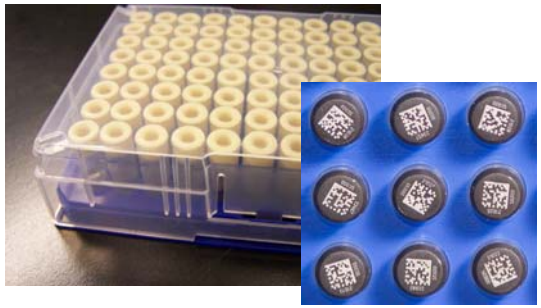
Liquid Culture

On-site archival sample
(barcode-labeled plates)

Off-site archival sample
(barcode-labeled plates)

DNA prep for sequencing

Sample Storage Production



Working sample
(barcode-labeled tubes)

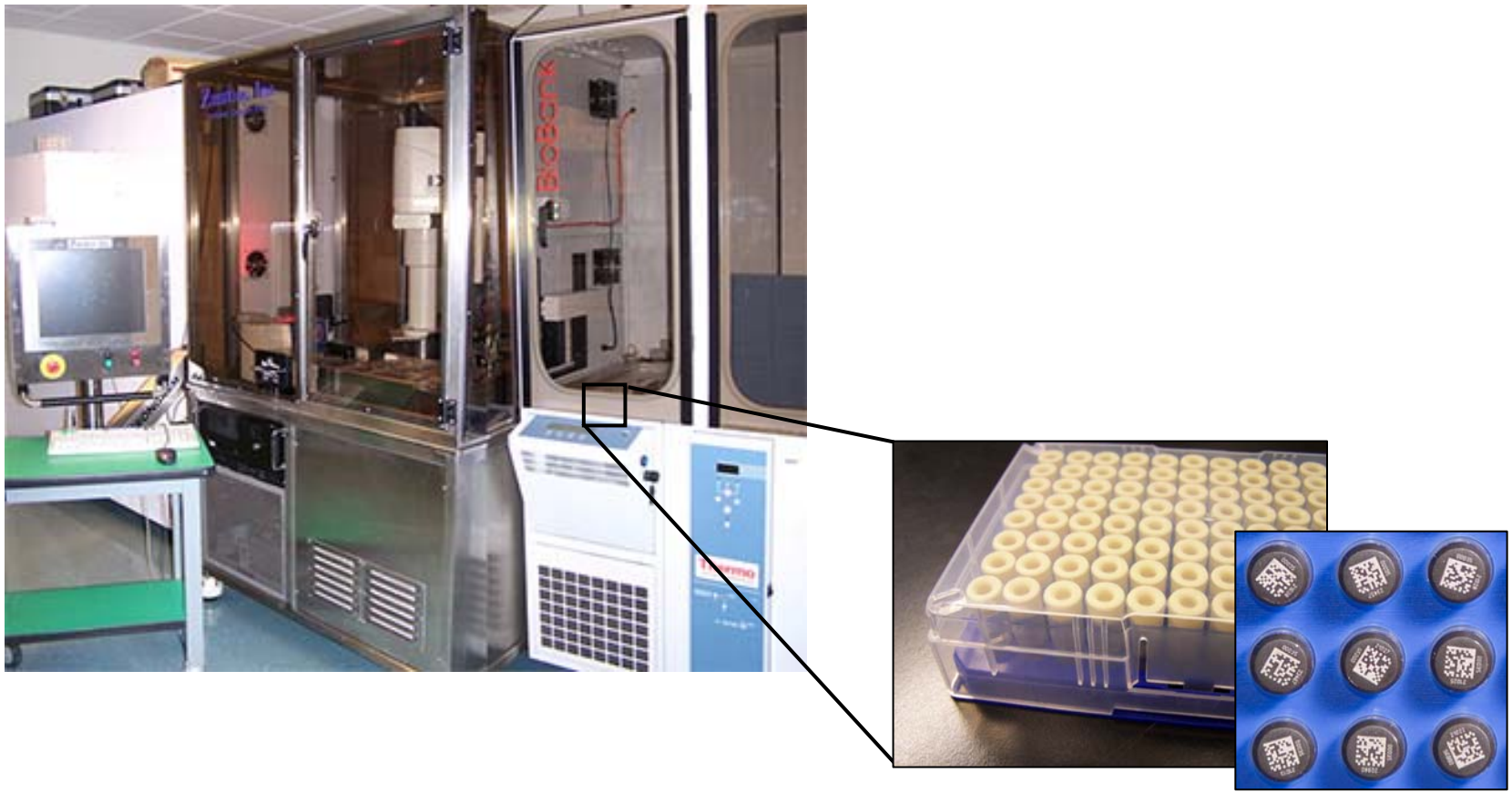
Liquid Culture

On-site archival sample
(barcode-labeled plates)

Off-site archival sample
(barcode-labeled plates)

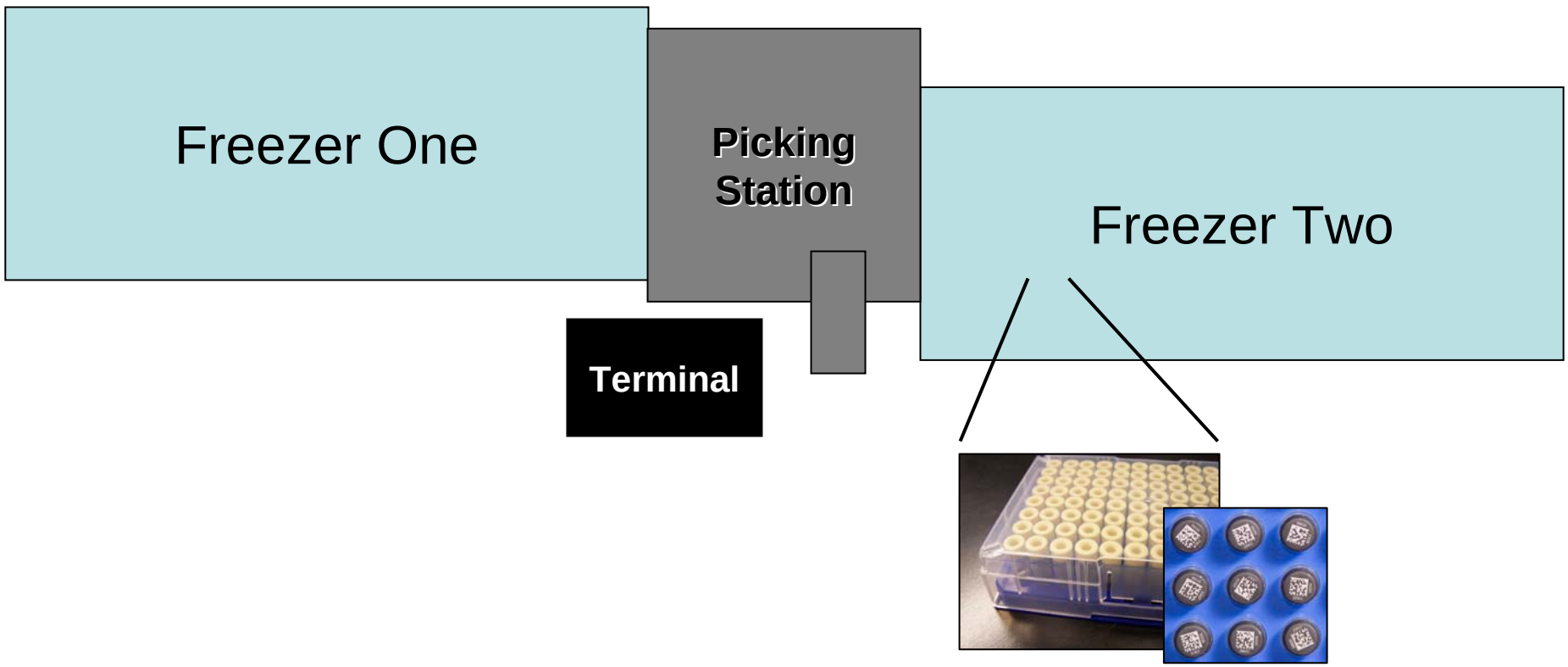
DNA prep for sequencing

Sample Storage—BioBank
Automated Freezer Storage System

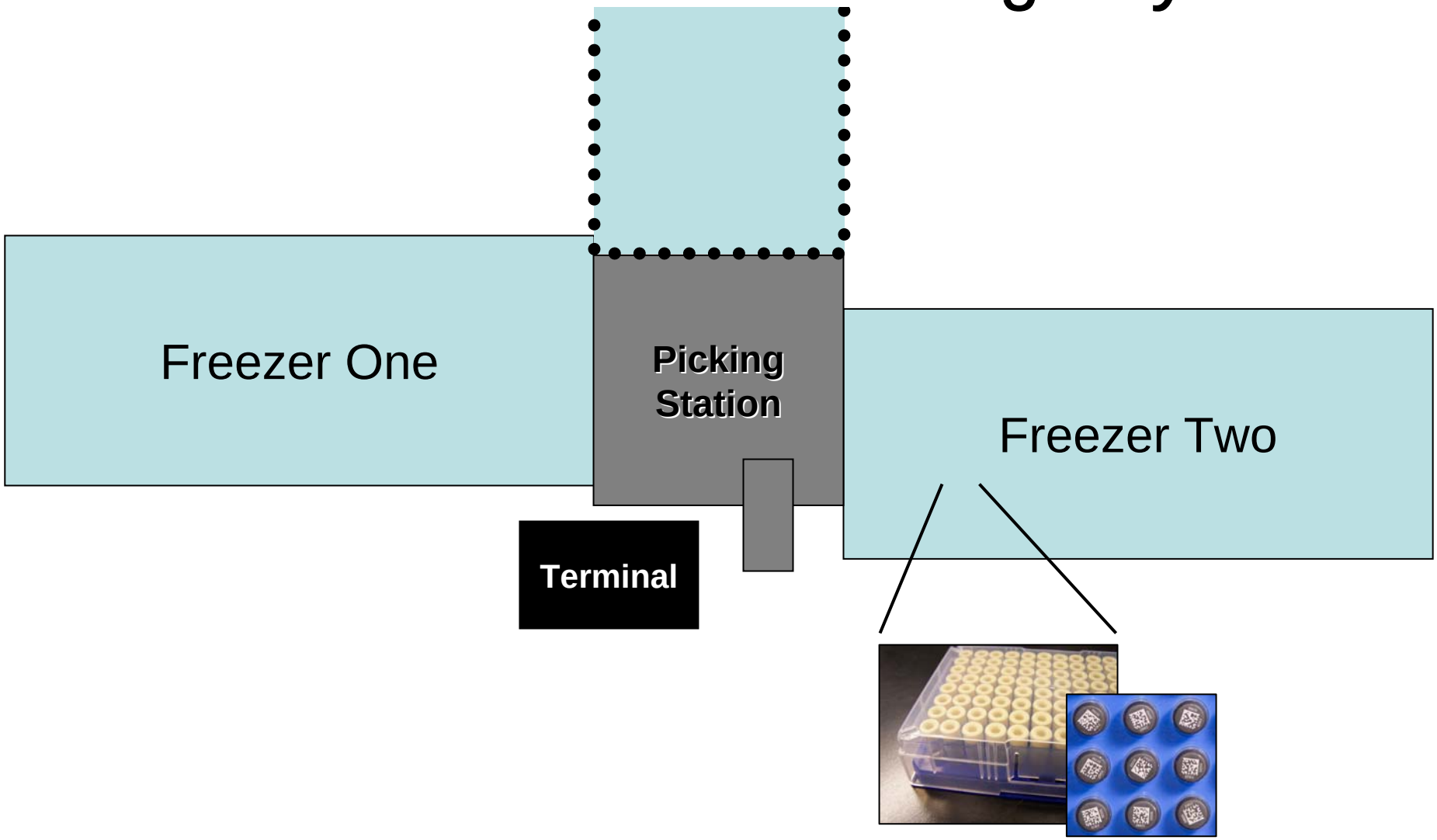


(**Thermo**, Zmation)

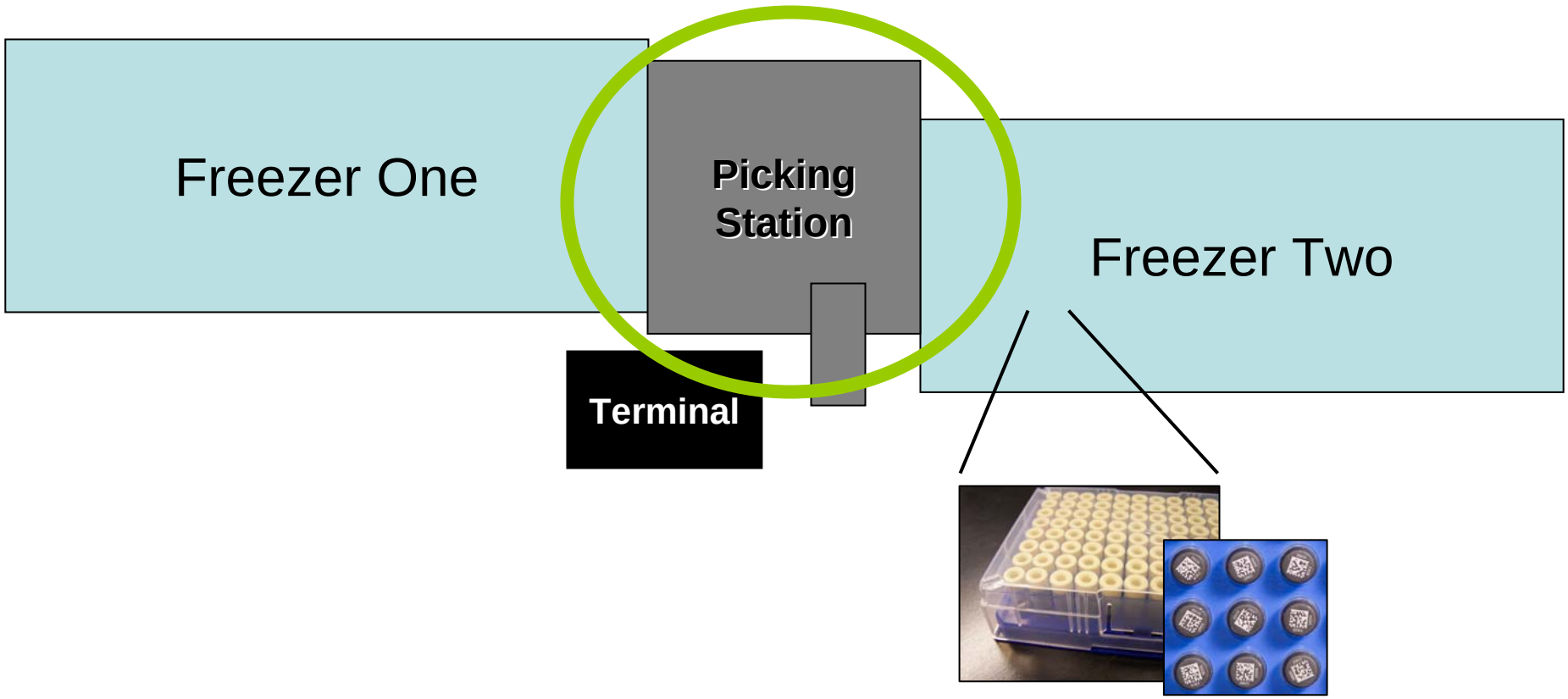
Sample Storage—BioBank
Automated Freezer Storage System



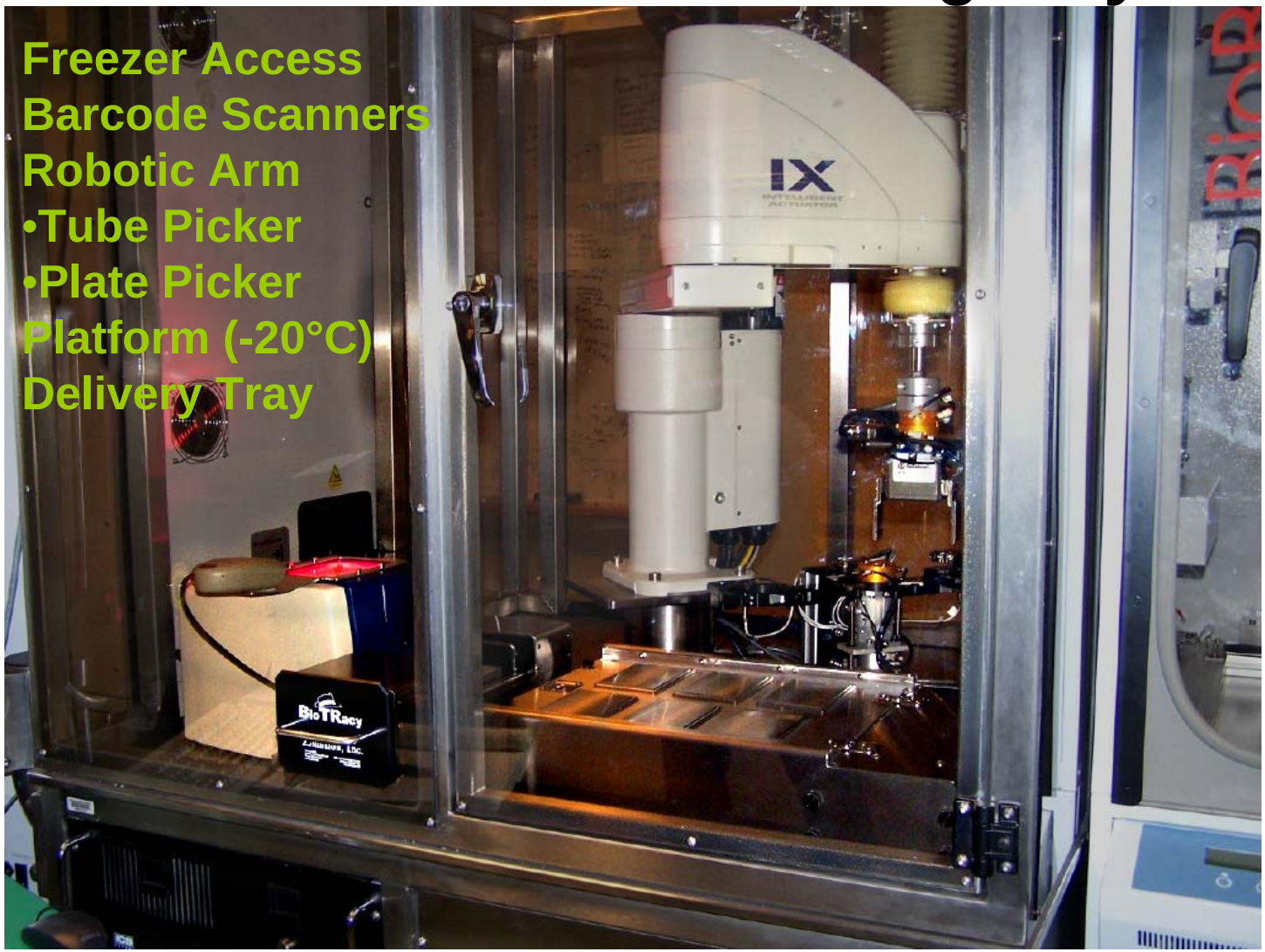
Sample Storage—BioBank
Automated Freezer Storage System



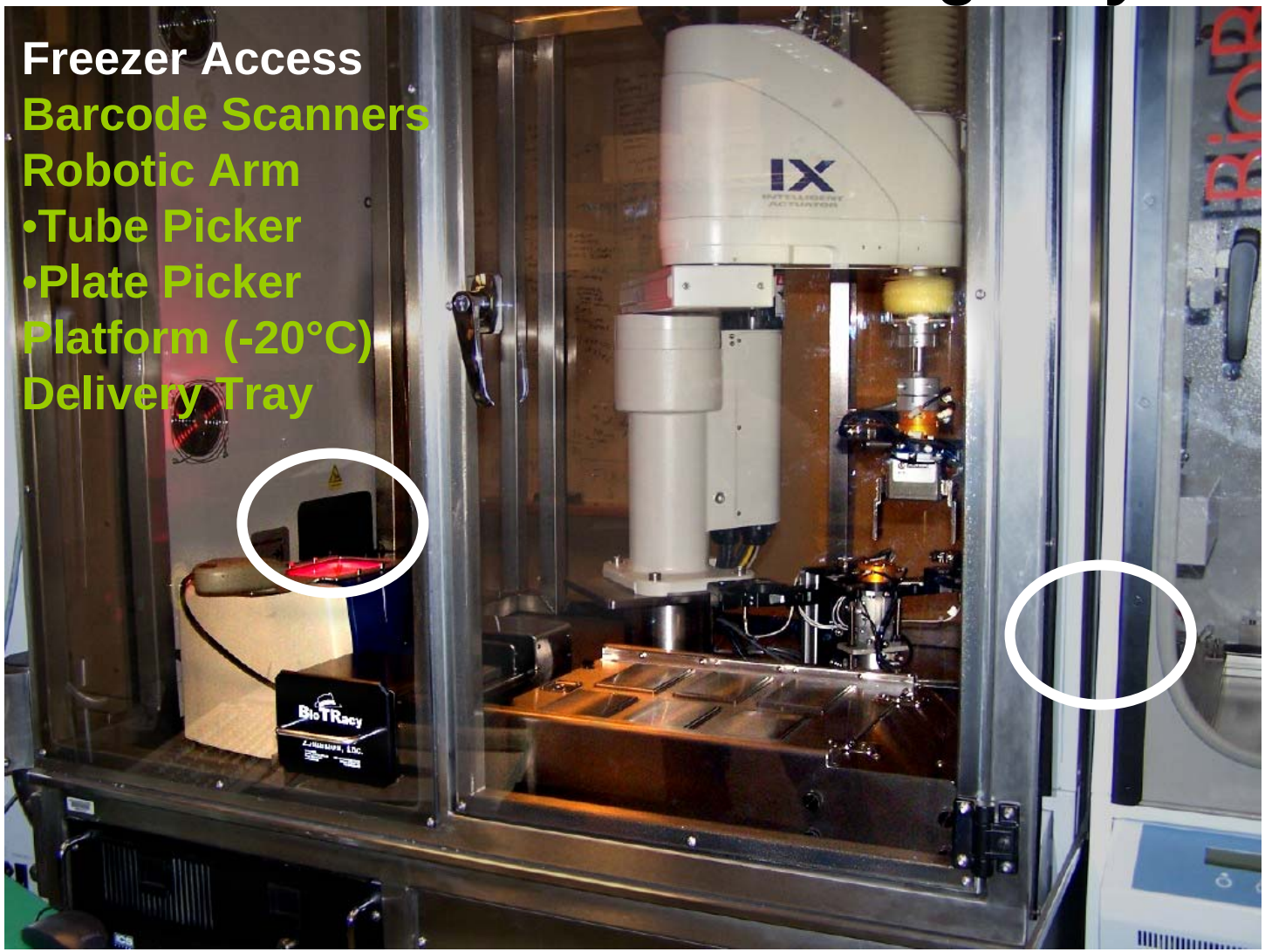
Sample Storage—BioBank
Automated Freezer Storage System



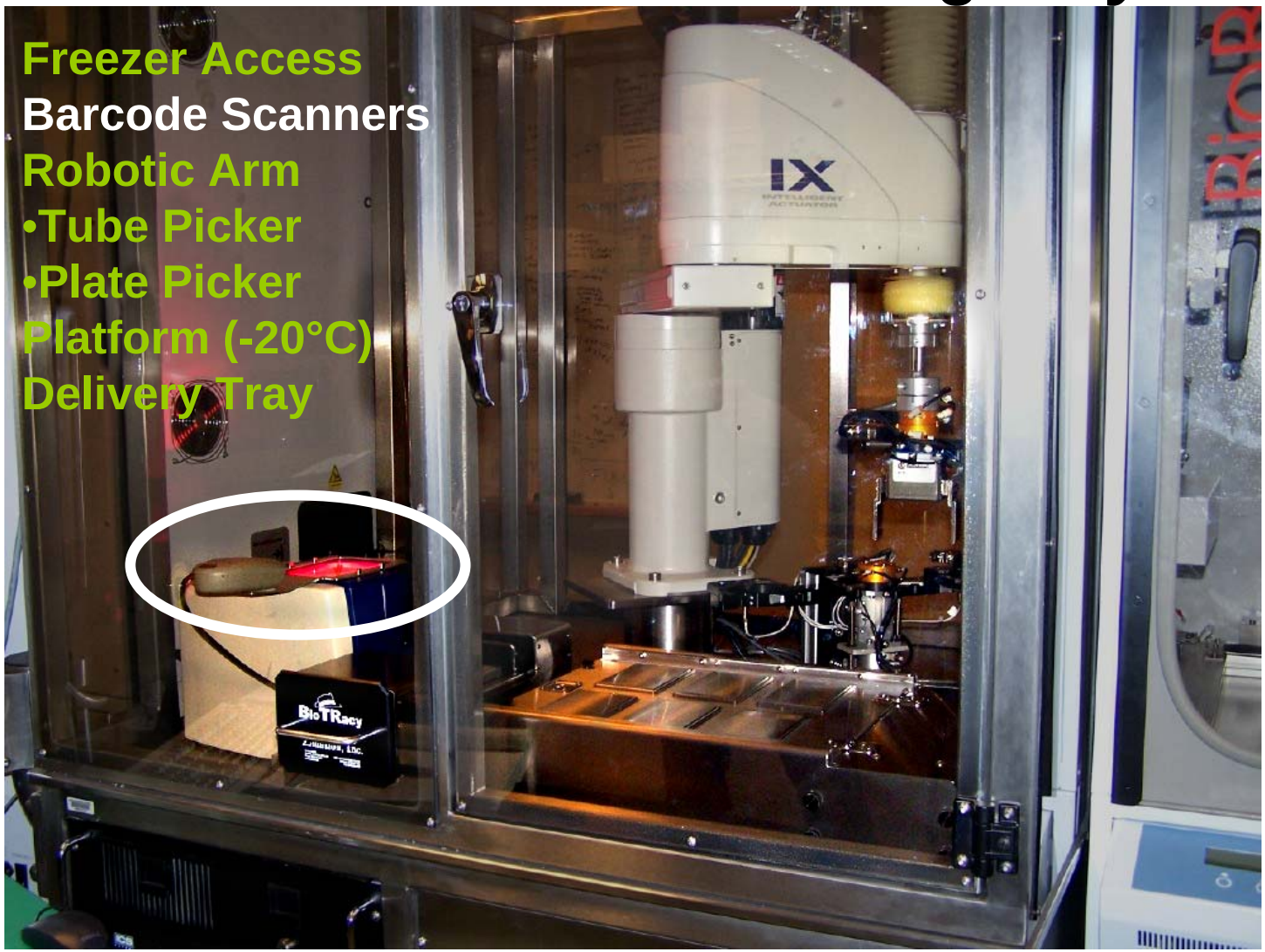
Sample Storage—BioBank
Automated Freezer Storage System



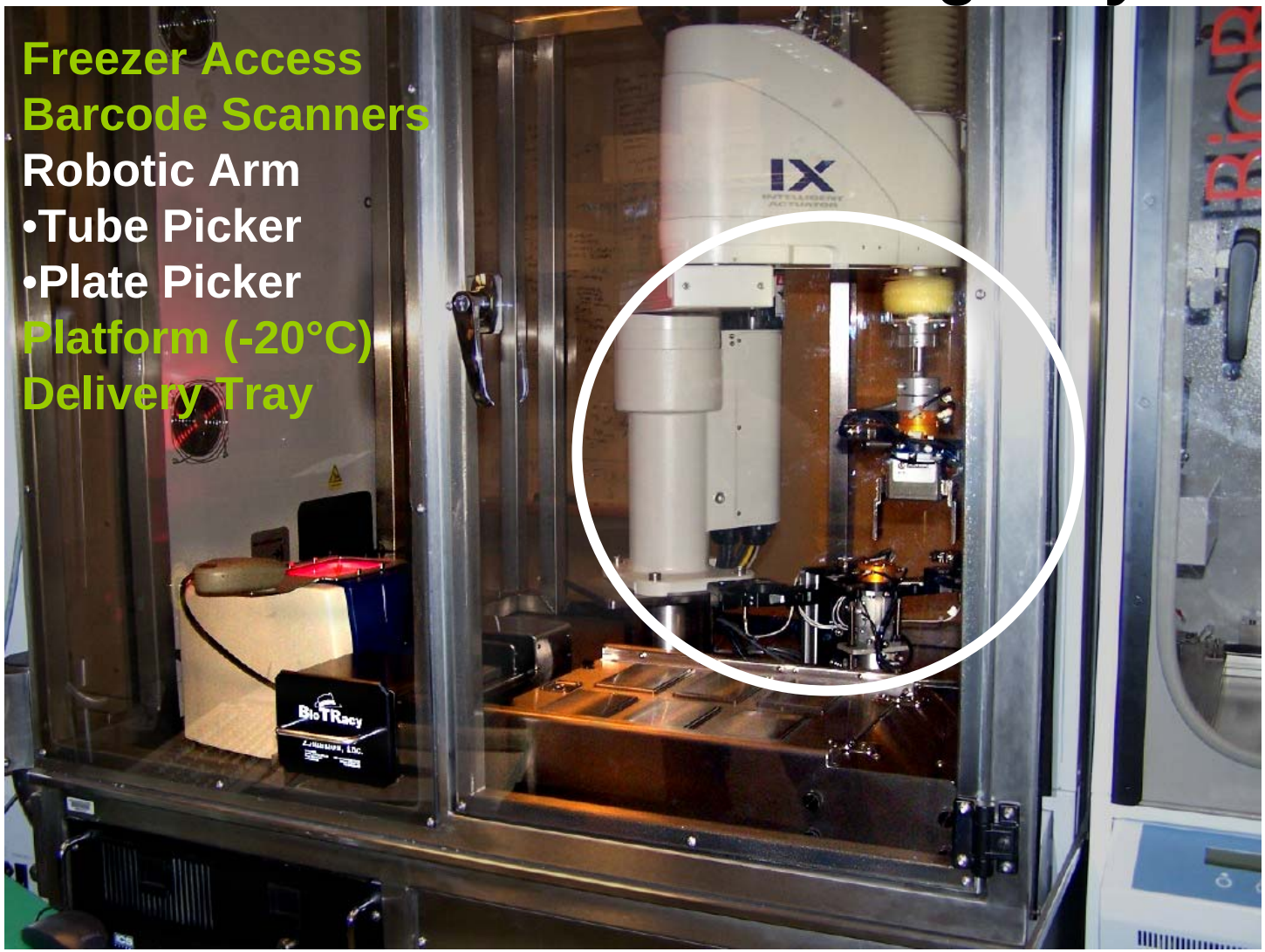
Sample Storage—BioBank
Automated Freezer Storage System



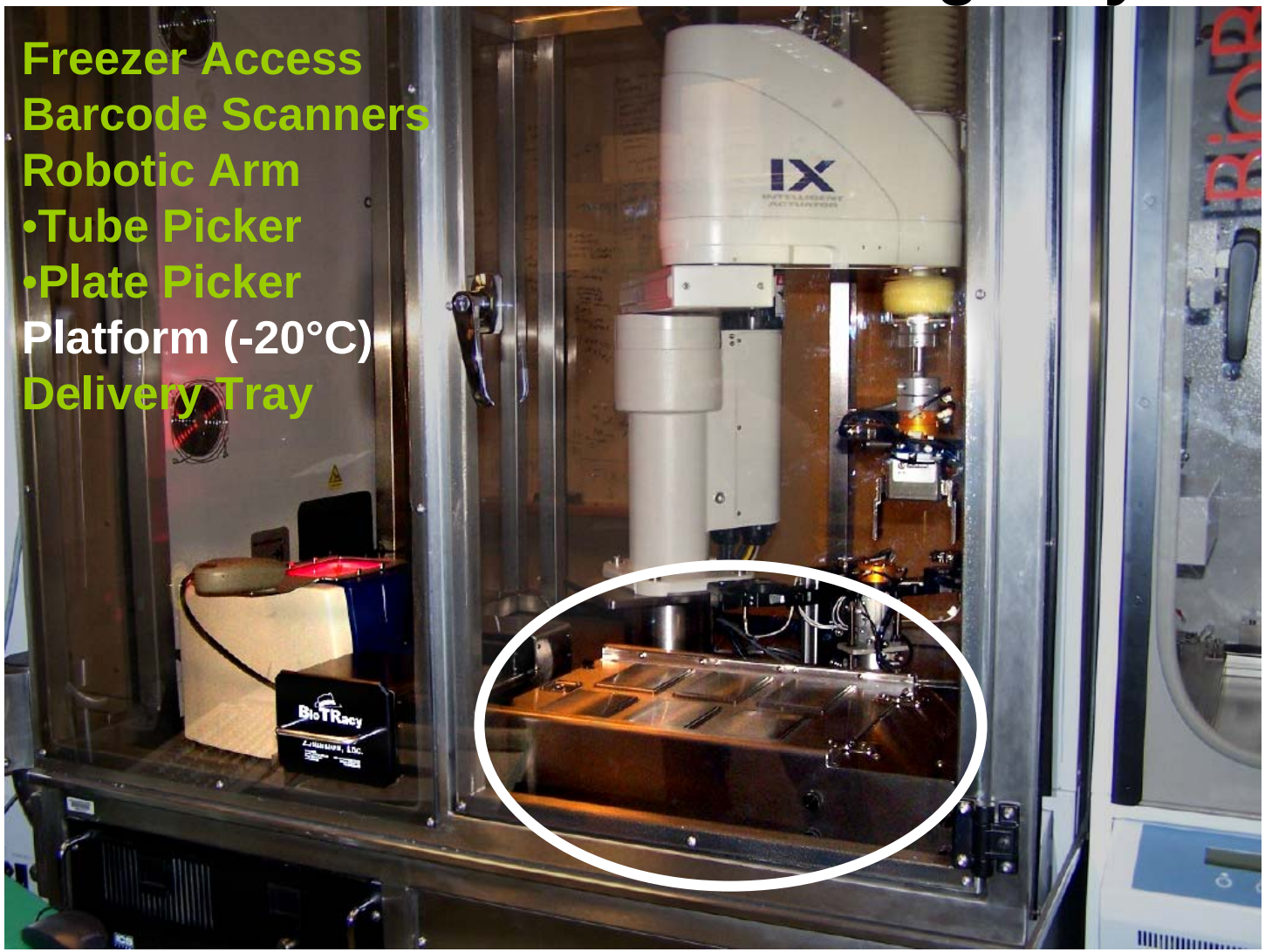
Sample Storage—BioBank
Automated Freezer Storage System



Sample Storage—BioBank
Automated Freezer Storage System

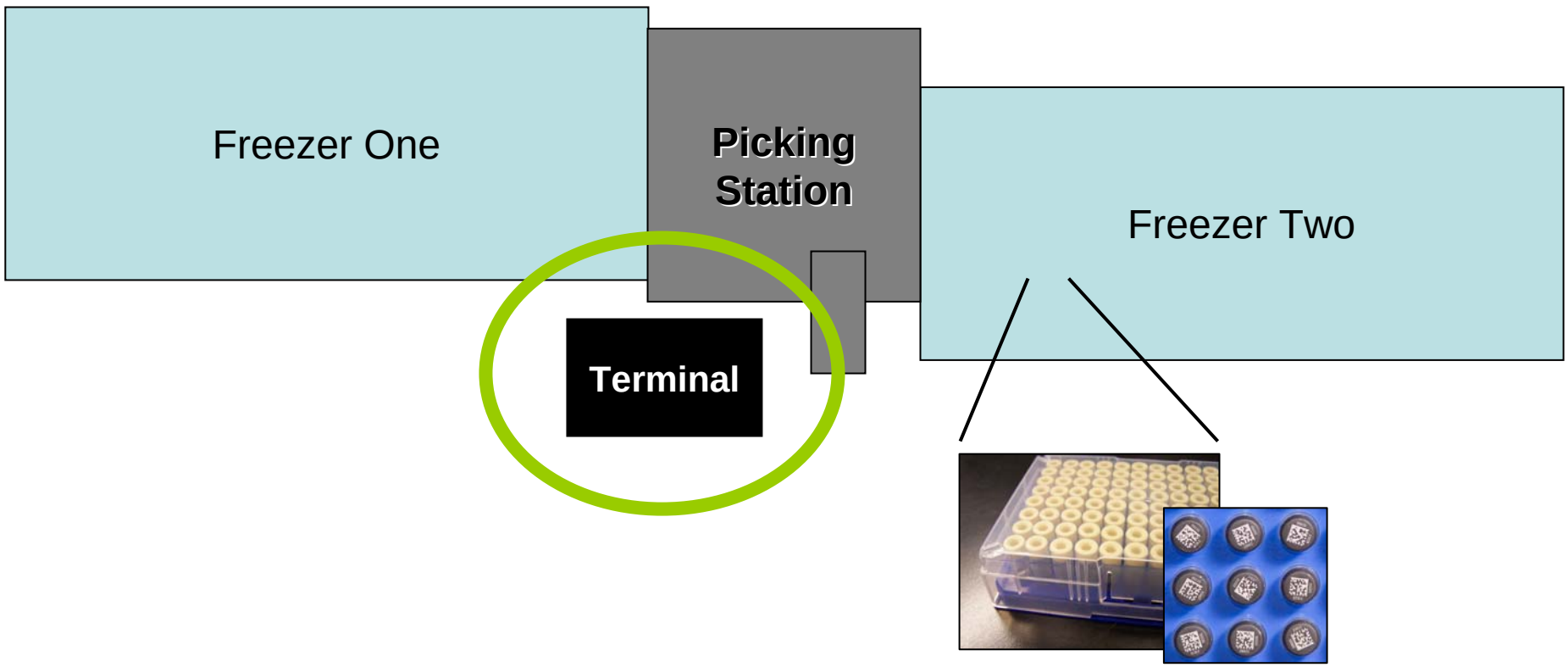


Sample Storage—BioBank
Automated Freezer Storage System



Sample Storage—BioBank
Automated Freezer Storage System

Sample Storage—BioBank
Automated Freezer Storage System

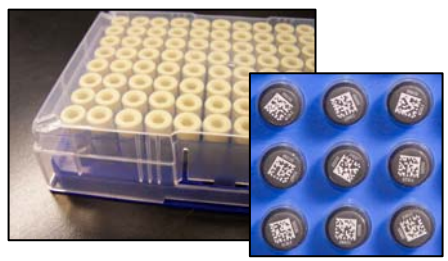


Sample Storage—BioBank
Automated Freezer Storage System

BioBank Software
- **Links Tubes, Plate Racks to Locations**
- **Retrieval of Individual Samples**
  - **Most efficient array format**
  - **User-controlled array format**
- **Flags samples after 6 freeze/thaws**
- **Flags samples after 5 years in storage**
- **‘Domain’ Organization in Freezers**

Freezer One

Terminal

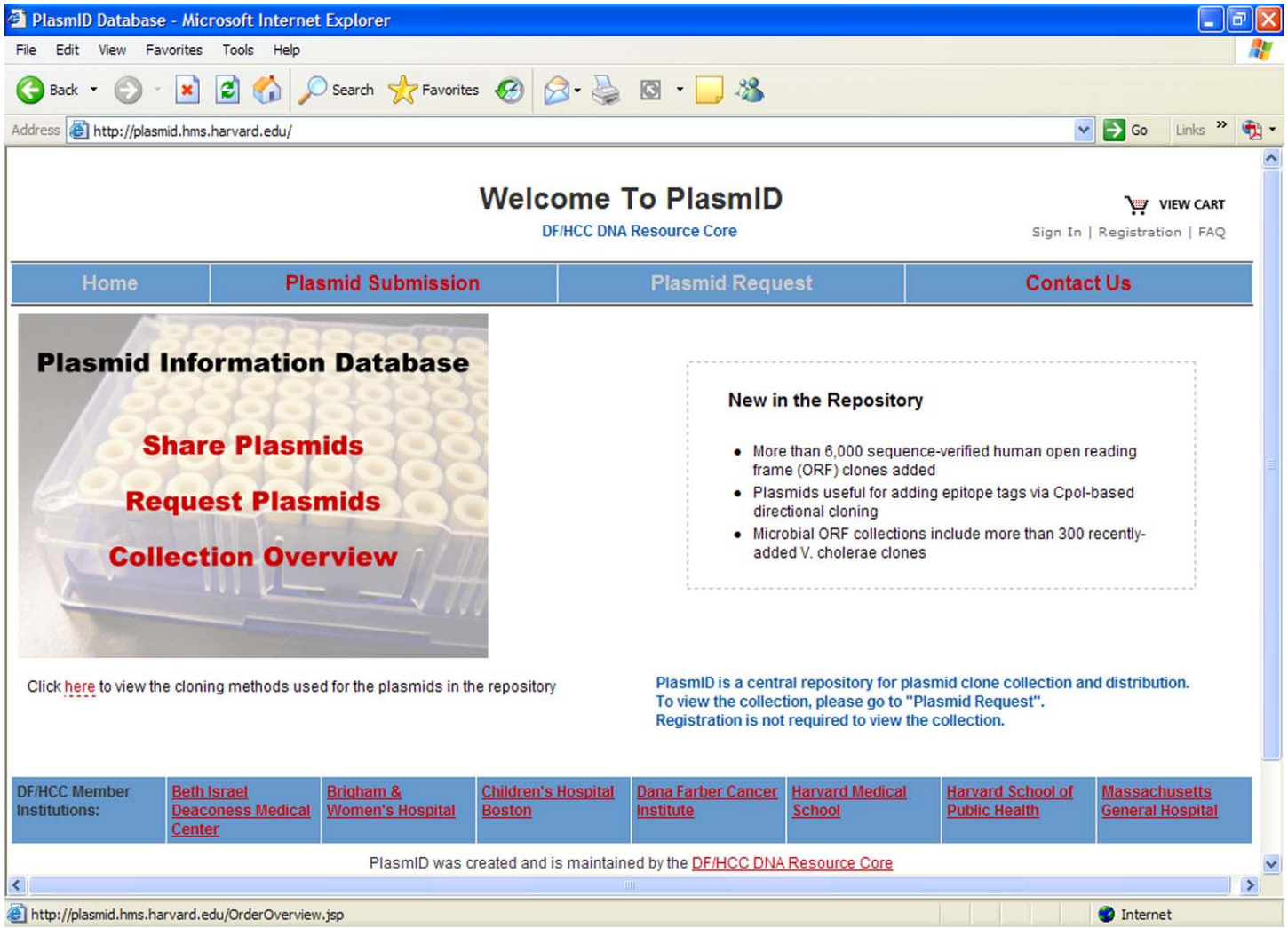


Connect samples to biological information about clones?

Plasmid Information Database
(PlasmID)

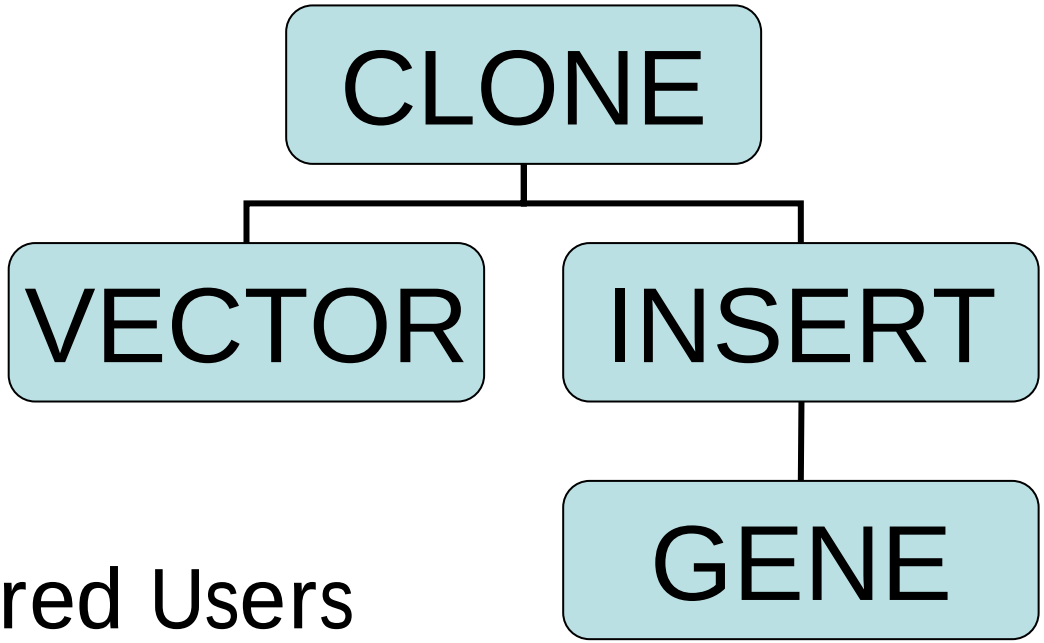
(Zhou et al. 2006 NAR)

Plasmid Information Database (PlasmID)



http://plasmid.hms.harvard.edu

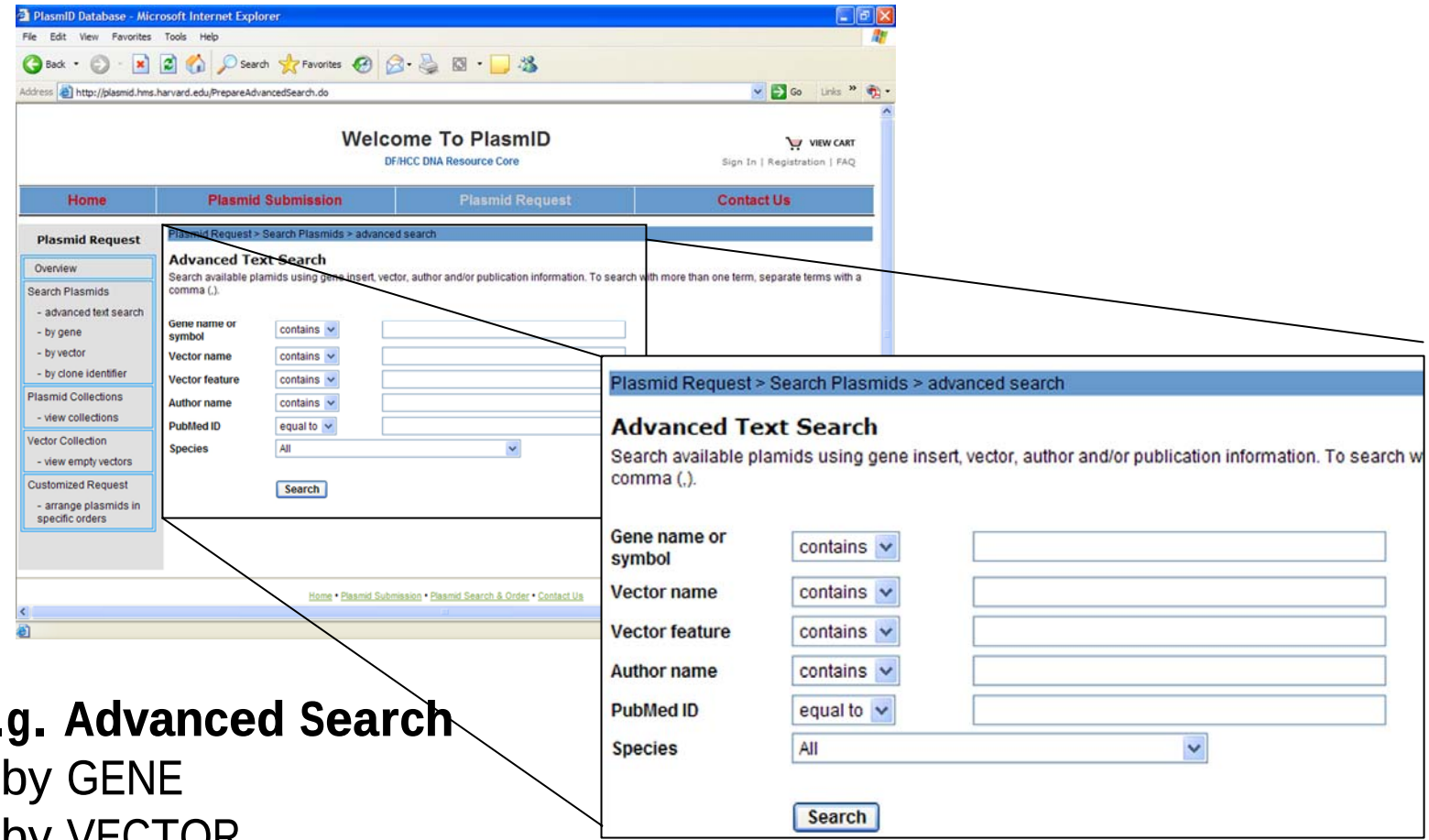
PlasmID



Also:

- Registered Users
- Request Histories
- Special MTAs, Restrictions
- Sample Storage Locations

Organization Facilitates Searches



e.g. Advanced Search

- by GENE
- by VECTOR
- by AUTHOR
- by PUBLICATION
- optional limit to SPECIES

Clone Collections—Issues

Shared Repositories at HIP

DF/HCC DNA Resource Core
~40,000 clones+

PSI-MR
~65,000 clones+

FLEXGene

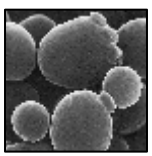
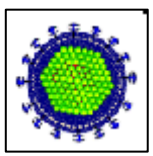
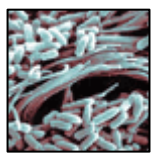
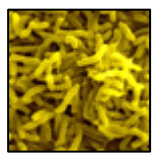
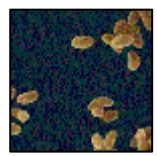
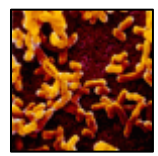
- ORF clones
- Human, yeast & Prokaryotes
- Genome-scale collections
- Full-length ORFs

Others

- ‘Empty’ vectors molecular techniques
- Many organisms
- shRNAs, genomic fragments, cDNAs
- Full-length, partial and mutant forms

PSI Labs

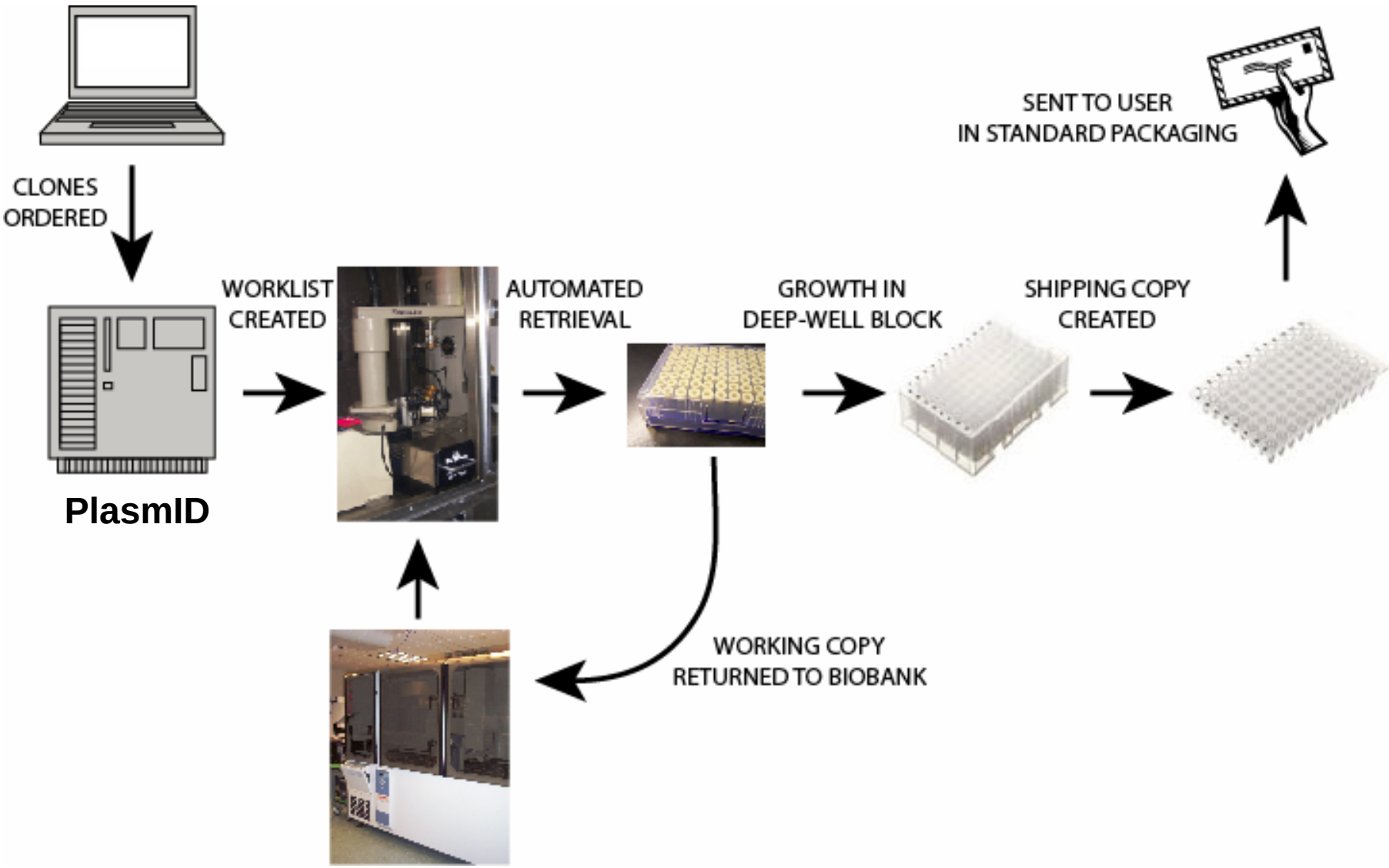
- Expression vectors (various systems)
- Expression clones
- Many organisms
- Full-length, partial and mutant ORFs


Clone Request Fulfillment

- Minimally restrictive MTAs
- Pre-approval agreements for MTAs
- Accommodate special MTAs, restrictions
- On-line request via PlasmID

Clone Request Fulfillment



FLEXGene Clones

CLONE PRODUCTION	SEQUENCE VALIDATION	CLONE STORAGE	CLONE DISTRIBUTION
Informatics			
FLEXGene LIMS	ACE software	BioBank, PlasmID	PlasmID
Automation			
Liquid handling, Colony picking	ABI 3730xl	BioBank	BioBank

Institute of Proteomics

Harvard Medical School

Joshua LaBaer, MD, PhD, Director

Jason Kramer, Manager
DF/HCC DNA Resource Core

Dongmei Zuo, Yanhui Hu, Michael Collins—HIP Bioinformatics & IT
Mauricio Fernandez—Robotics Engineering

FUNDING SOURCES: **DF/HCC, NIGMS at NIH (PSI)**

DF/HCC

PSI Protein Structure Initiative