

...TASTTTTTTCTARTGATCGGCGAC-
...GAAAGGGGCGTTCCAGCGGGACACACTATCCG...
...GTGAGGATGCGATTIGRGGGGGGCGAGGGTGORGGT...
...GTGGGCGCAGGEGGGGAGGAGGAGGGGGGGGCTTGGCGGAG...
...GATOTCGAGGGGOTCGGATGAGCAGGGGTTGGCTGGGCCAG...
...GTARAGAAAGGGGGGAGGGGTTACCGAGGGGGGGAGGGGTTAGTART...
...TTORAGGGGGGGGATGAGGGGGGATTGGCGGGACGGGGGGGGGATCOCT...
...ATARCGAGGTTGAGGGGTTCTACGGGAGGGGGGGCGGGGTCGGGGTCTAGGCT...
...ATATATGAGGAGGGGTOAGCATGAGGGGCGCTGGGGCGGGGGTGGGGTGGGAGT...
...CTOOCATGGGGTGGATCCACATGAGGGGGGGGAGGGGCGGGGGCGAGGGGGGGGGGG...
...CTOOGGAGGGTGGCTCCAGCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG...
...GGGGAGGGGGGGTCCAGGG...
...CGAGGGTGGTCCGGTGGTGGATTTGGGGTCCGGGGGGGGGGGGGGGGGGGGGGGGGG...
...CTOCCGCTGGGGAGGG...
...OTTCAGGGGGGGGACTGG...
...CTOCCAGGG...
...TGGGCTGG...
...GG...
...CGAGTGG...
...ACTTGG...
...ATTCTTGG...
...TGACTGG...
...TGAAGTGG...
...AAGAGGAGGG...
...TTTGGTAGGG...
...GGAGCGGG...
...ATGGGATTCGG...
...CTTACCGCTCTGG...
...GGGATGGGATTTCCGG...
...TTTACCGAARGACCGGG...
...TACTOCTATTAGCGAGGATTTCTGTAGCTOCTGATGCTAGGGGGGGGGGGGGGGGGGG...
...AGGGTCCATTAAAGTGAAAGGCTCAGCGAGGAGATGATGGGGGGGGGGGGGGGGGGGG...
...TTCAGGTTAGTTCTAAGGTAAGTACATAGATATGGGGGGGGGGGGGGGGGGGGGGGGGG...
...TATGAGGG...
...TTCAGGG...

Identifying Problems in a High-throughput Pipeline: a Failure Mode Analysis

George Yang
Production Coordinator, Sequencing
B.C. Genome Sciences Centre
Vancouver, British Columbia, Canada

**DNA Sequencing Satellite Meeting,
ABRF 2006**
Long Beach, California, USA
February 11, 2006

Background

- In the past decade, the demand for DNA sequence data has driven the transformation of sequencing from a research activity into a manufacturing process
- Basic manufacturing principles are becoming increasingly important in high throughput sequencing facilities where there is a constant drive to increase quality, increase efficiency, and decrease operating costs
- High throughput sequencing centers are focused on establishing automated procedures that will maintain long read lengths and high overall success rates
- The result is an exponential increase in the amount of quality sequence data generated from such facilities

Evaluating sequence data on a larger scale

- BCGSC has the capacity to generate 10,752 sequence reads per day (~3.9 million reads/year)
- It is neither practical nor economical to evaluate each and every sequence read daily
- Due to the large scale of sequence data generated , high throughput sequencing centers monitor sequencing success by referencing 1) overall pass rates 2) average read lengths
- The centers typically report failure rates on the order of about 10%, the causes of sporadic sequencing failures were seldom analyzed in detail and had not, in the past, been formally reported.

Goal

- Understand the baseline sequence failure rate that can be expected from our automated 384-well sequencing pipeline due to *process-related* causes. This is observed in the form of sporadic failed wells or “dropout” wells on a reaction plate
- To investigate and systematically identify the failure mode in each dropout well of a 384-well reaction plate
- Provide troubleshooting insight for similar high-throughput automated sequencing platforms
- Yang GS, Stott JM, Smailus D, Barber SA, Balasundaram M, Marra MA, Holt RA. **High-throughput sequencing: a failure mode analysis**. BMC Genomics. January 2005;6:2.

LIMS plate view: 384-well sequencing plate

Average Q20: 904bp

Pass rate: 97.4%

Runs 25867 25868 25869 25870 :

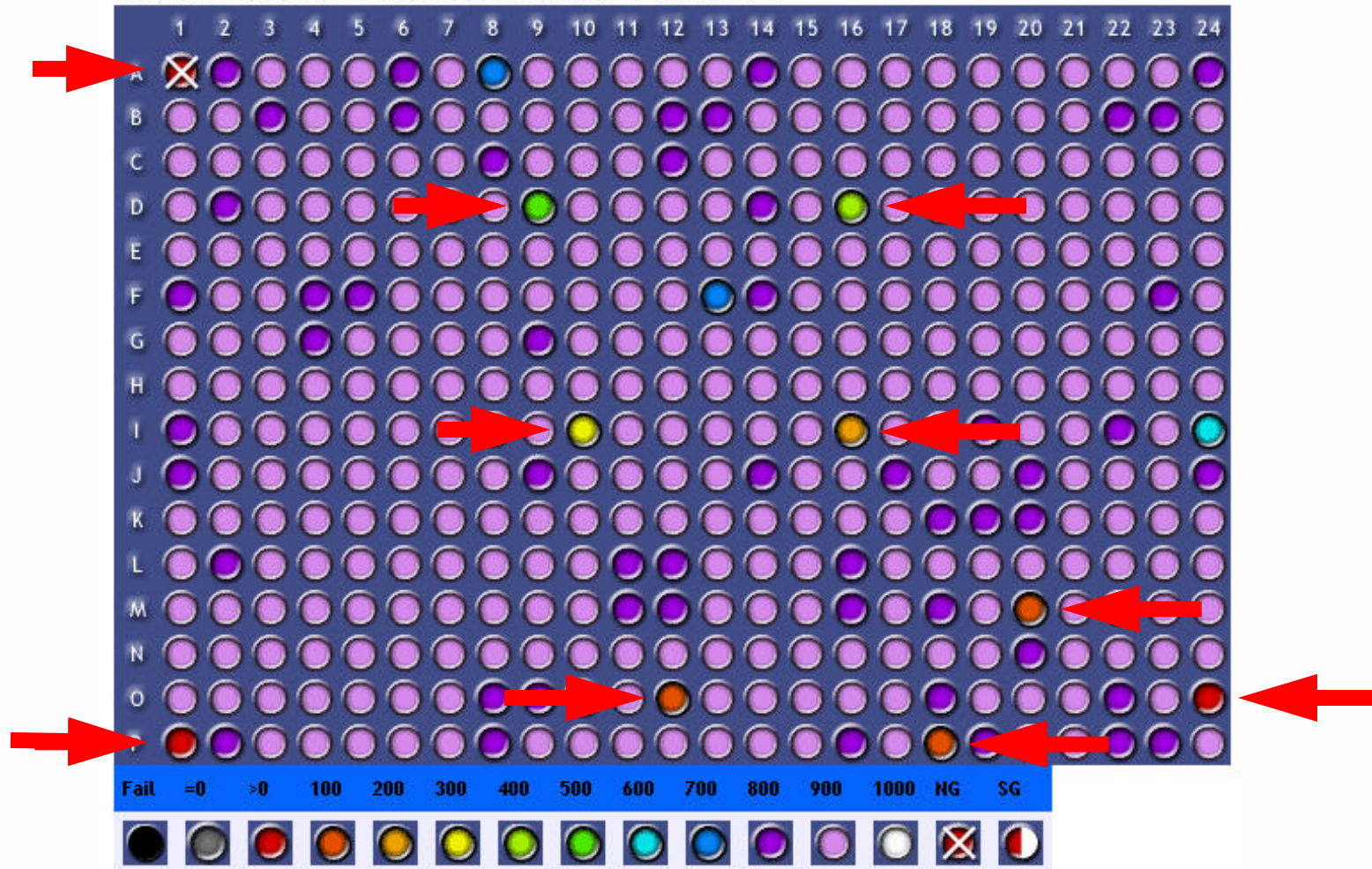


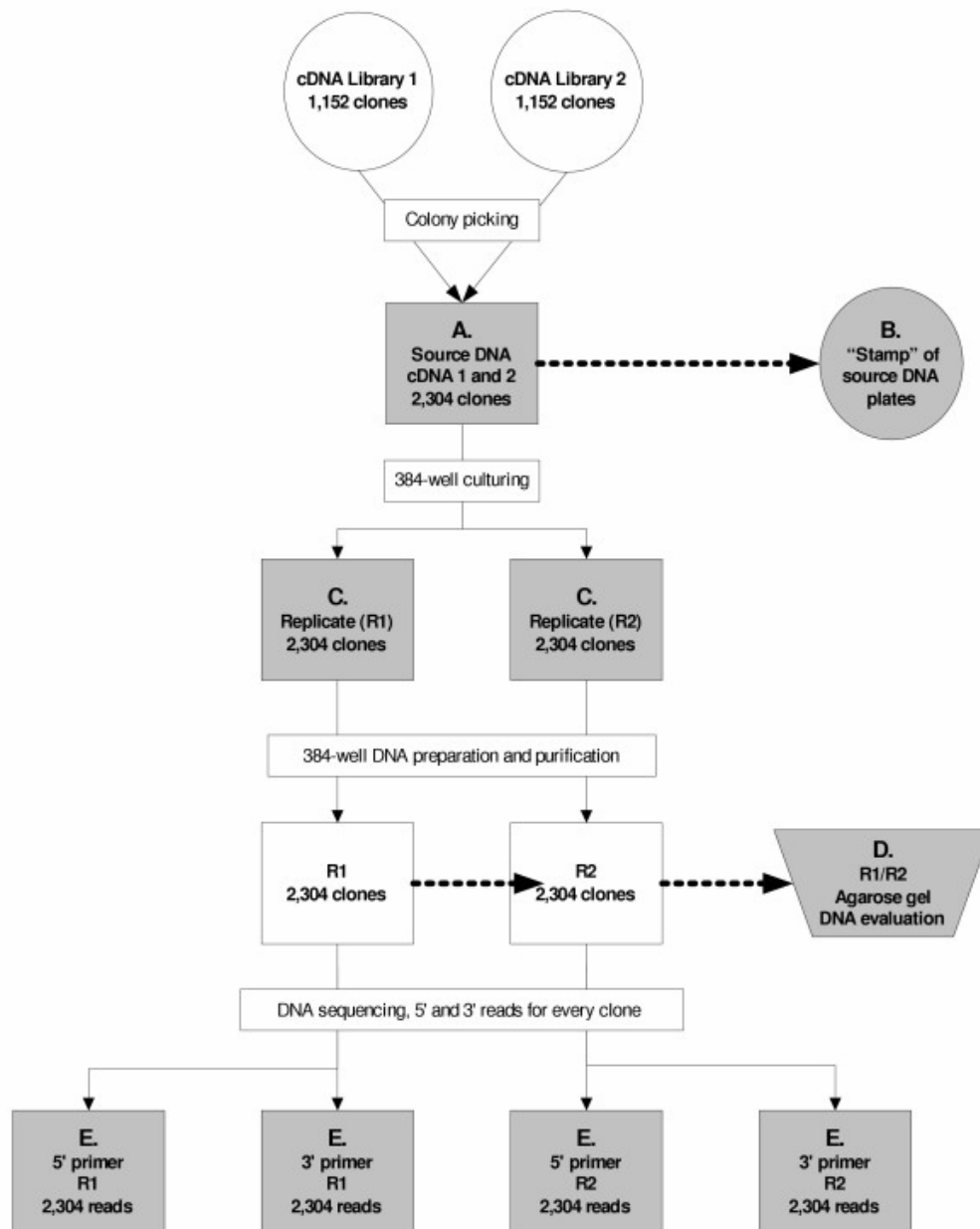
“Dropout” wells

Average Q20: 904bp

Pass rate: 97.4%

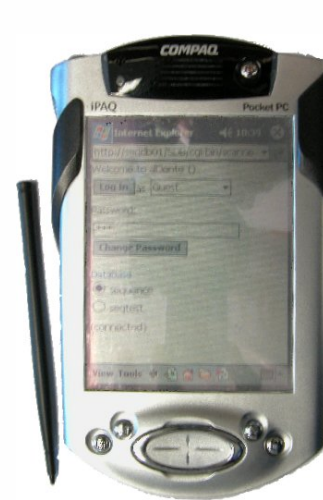
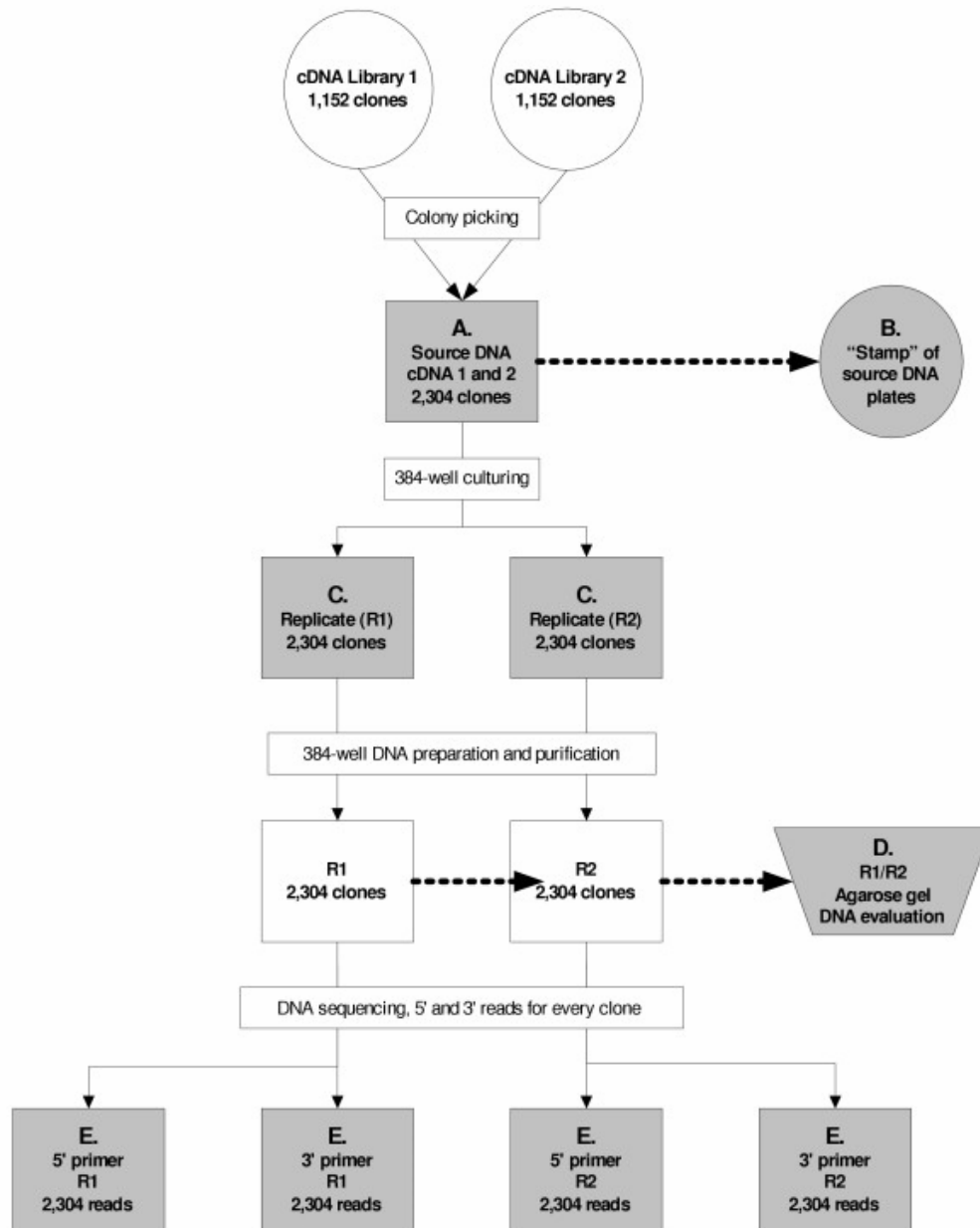
Runs 25867 25868 25869 25870 :



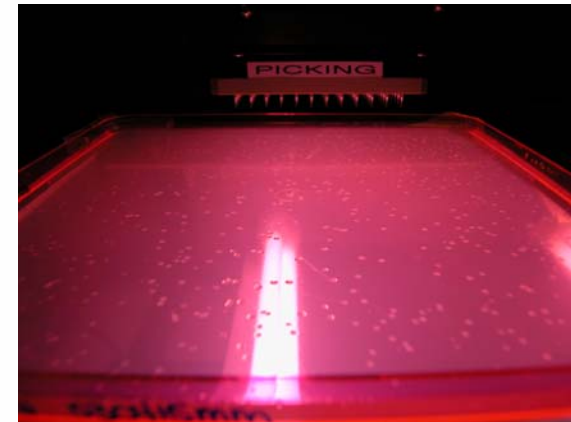
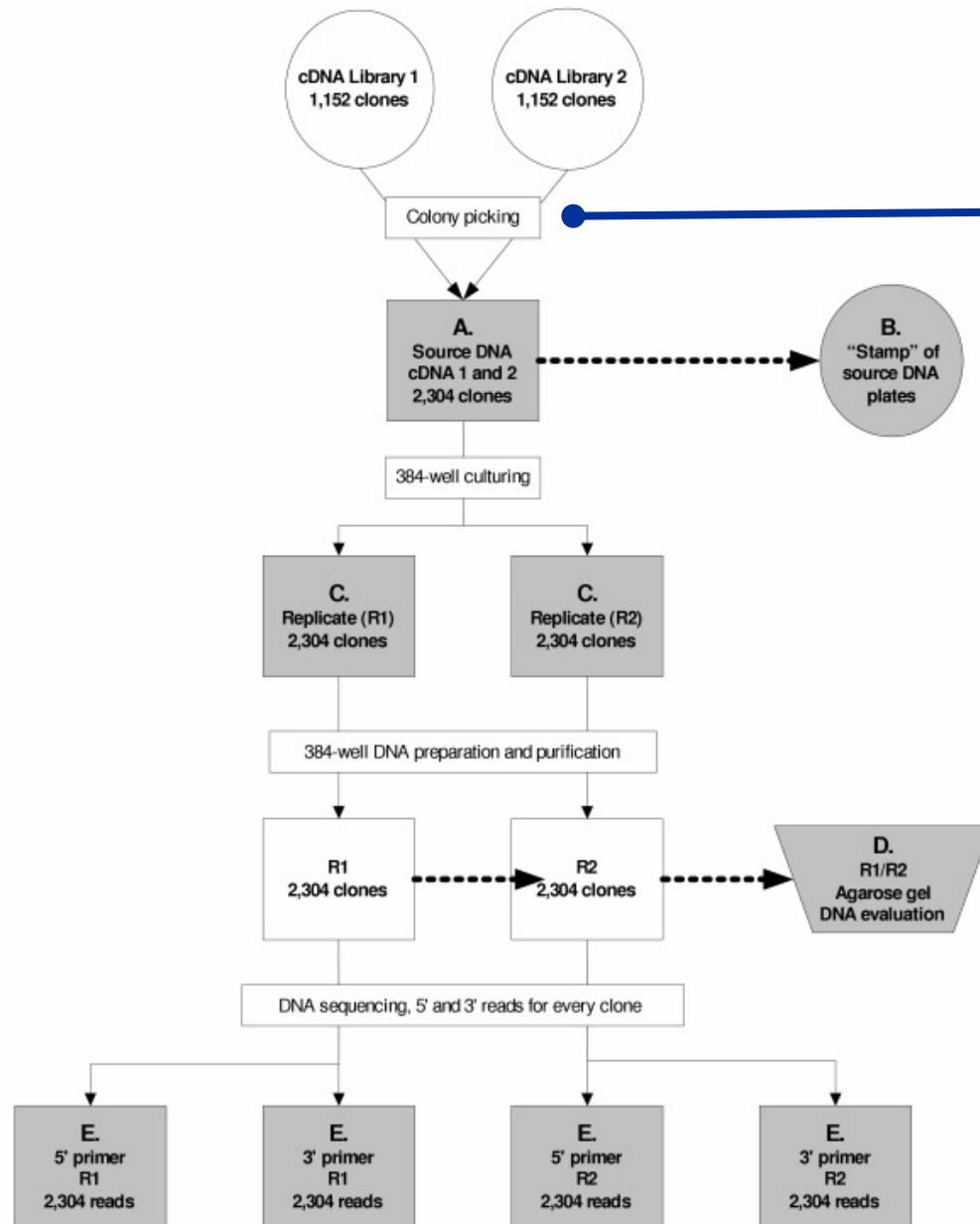


384-well automated sequencing pipeline

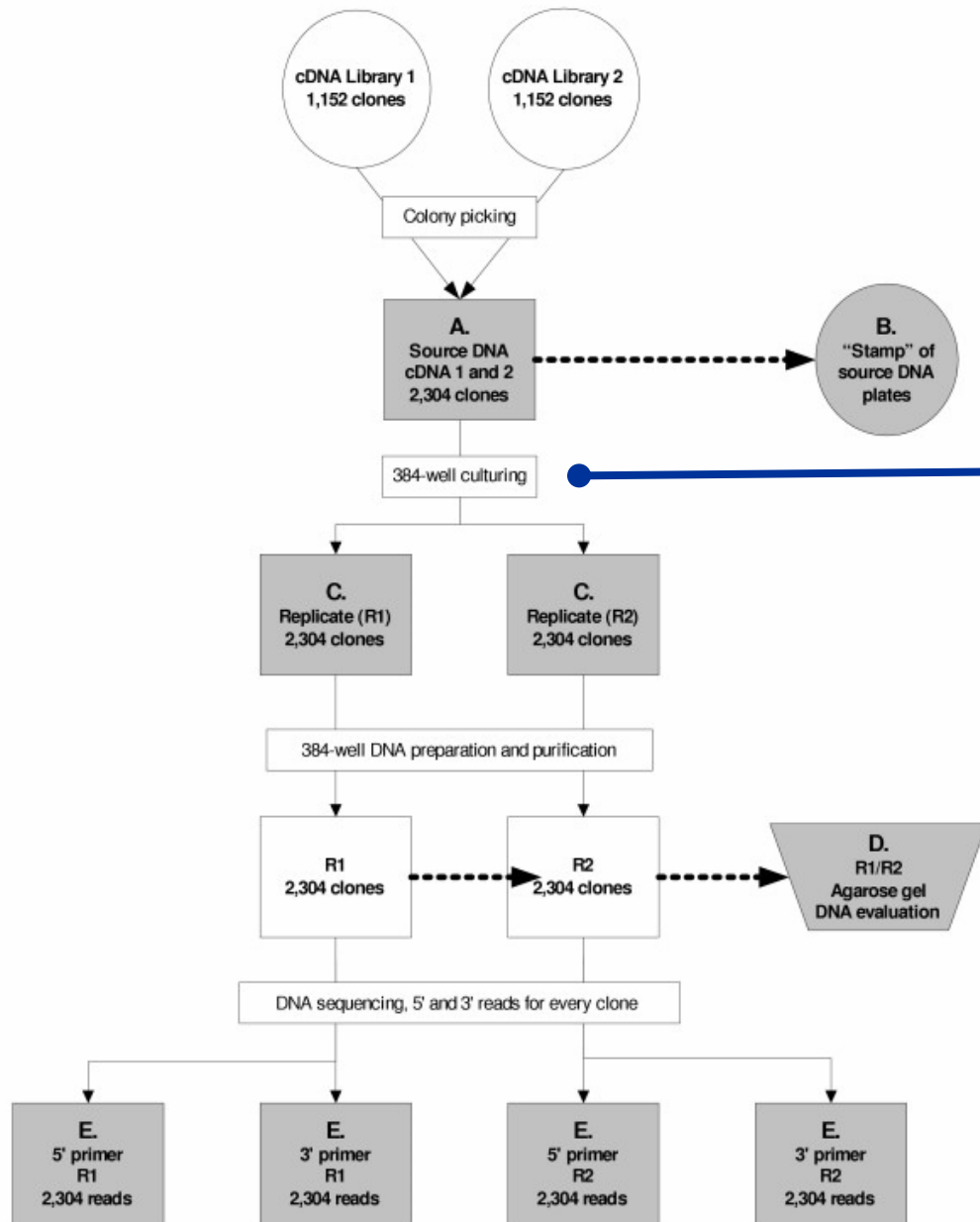
Barcoding / LIMS



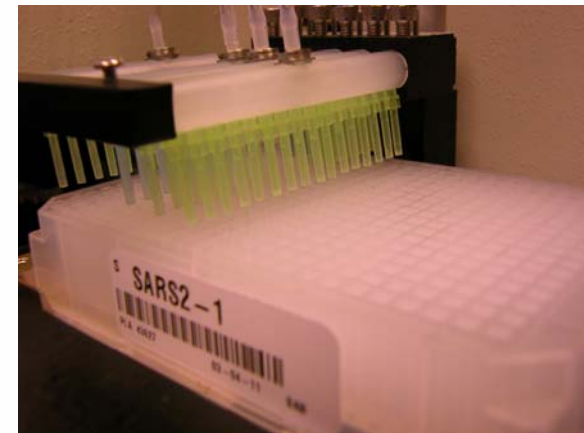
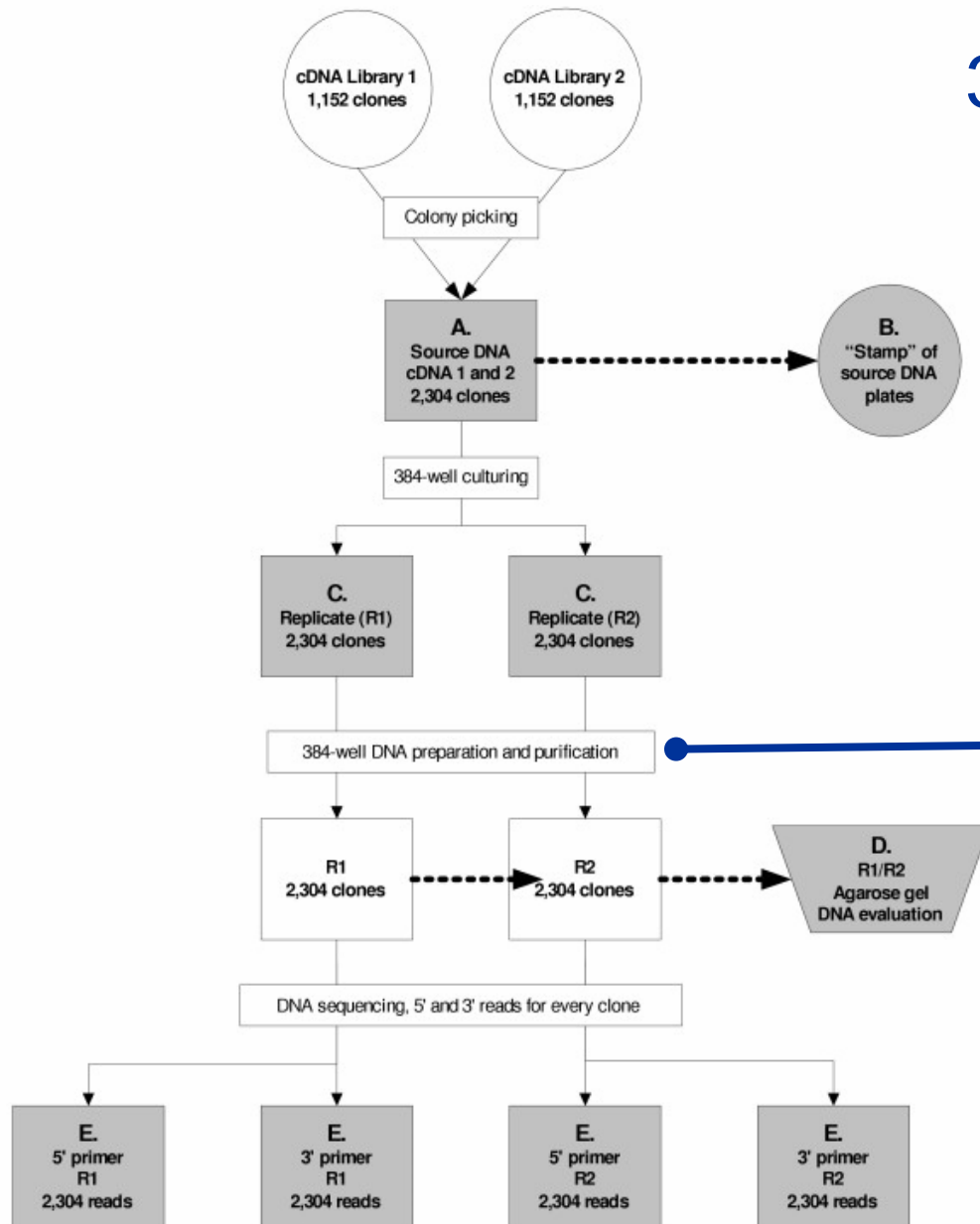
Colony picking



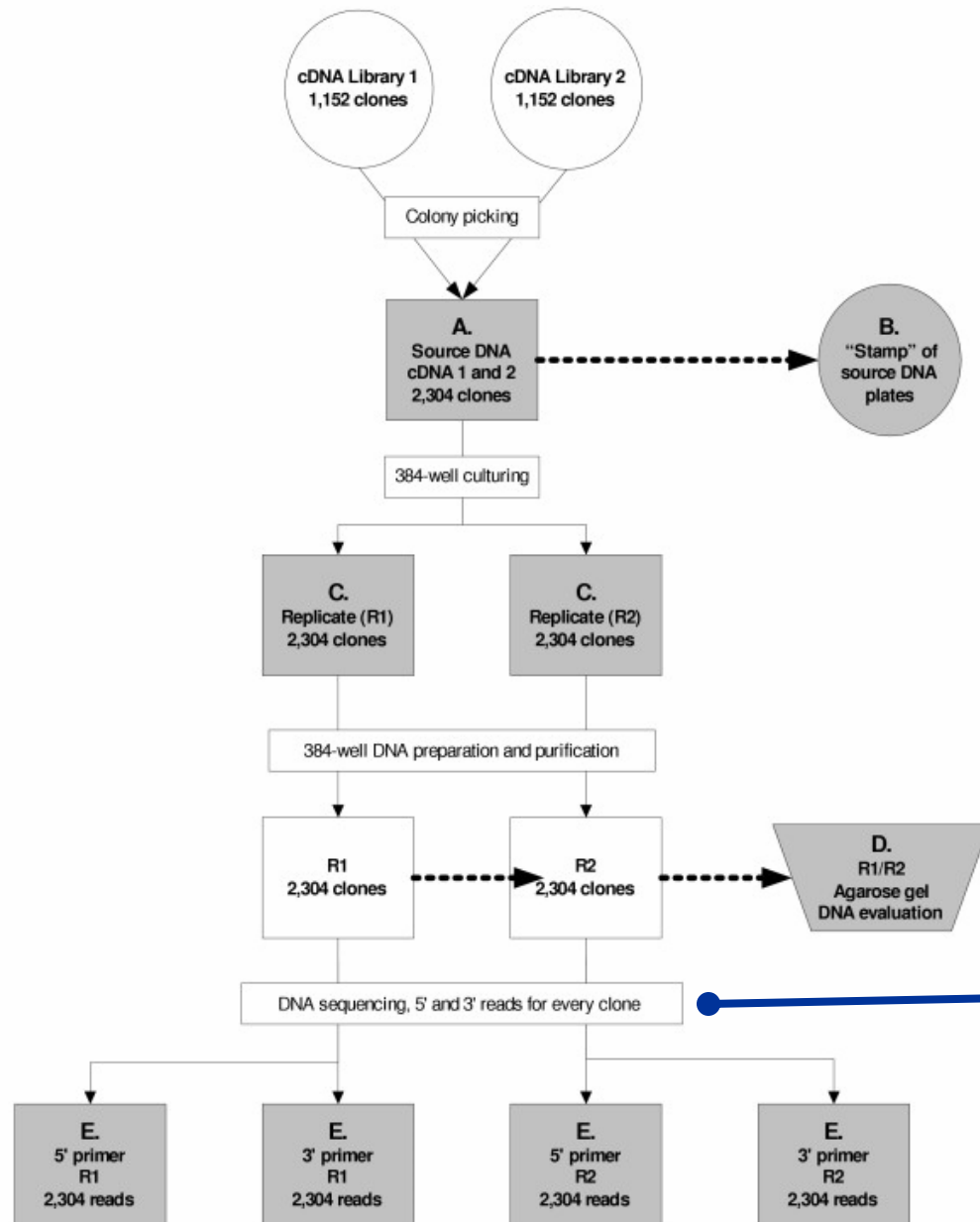
384-well bacterial culture setup



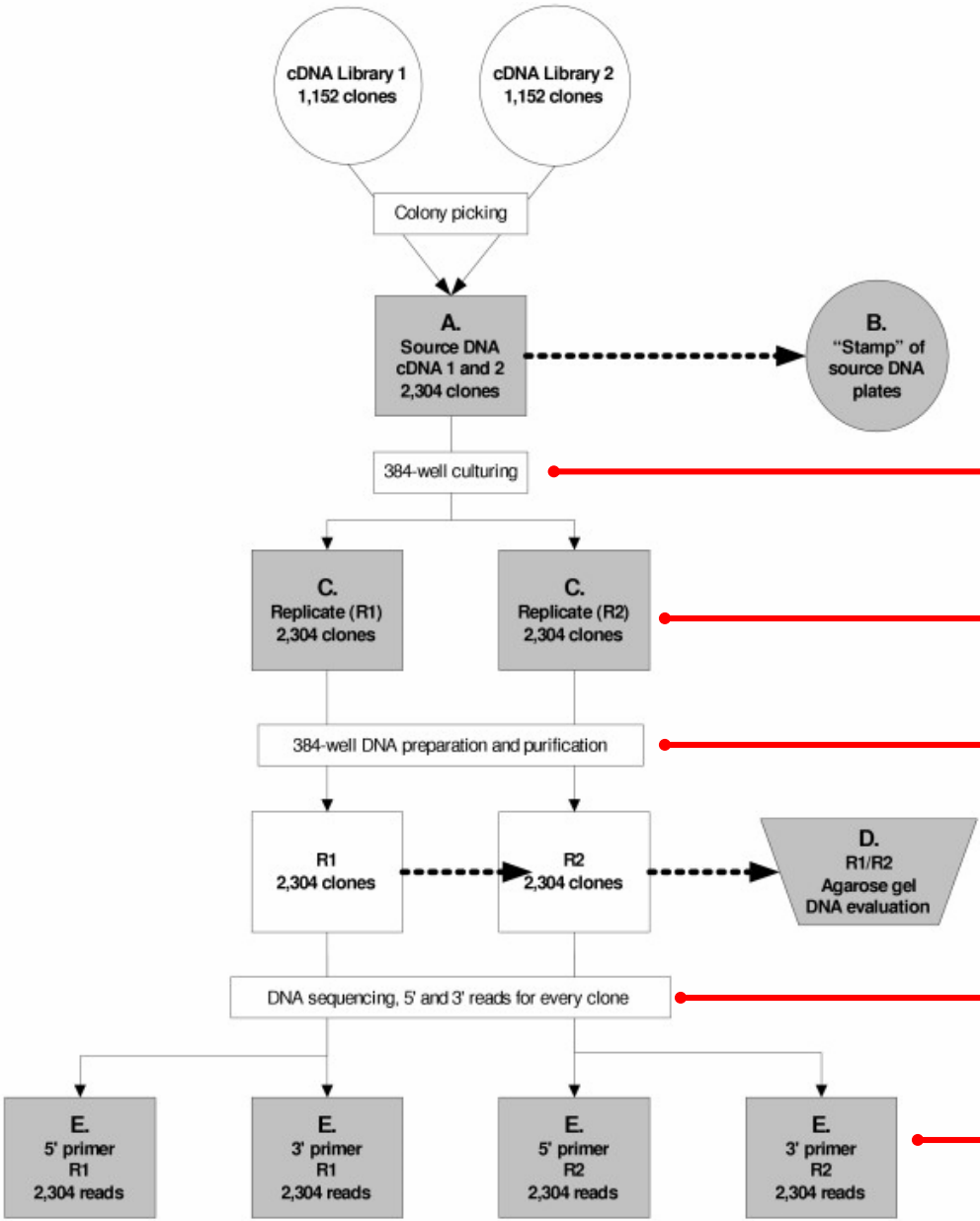
384-well DNA preparation



DNA sequencing



Experimental design



Bacterial growth assessment and replicate stamp onto agar

Bacterial growth assessment

Two replicates to verify process-related loss

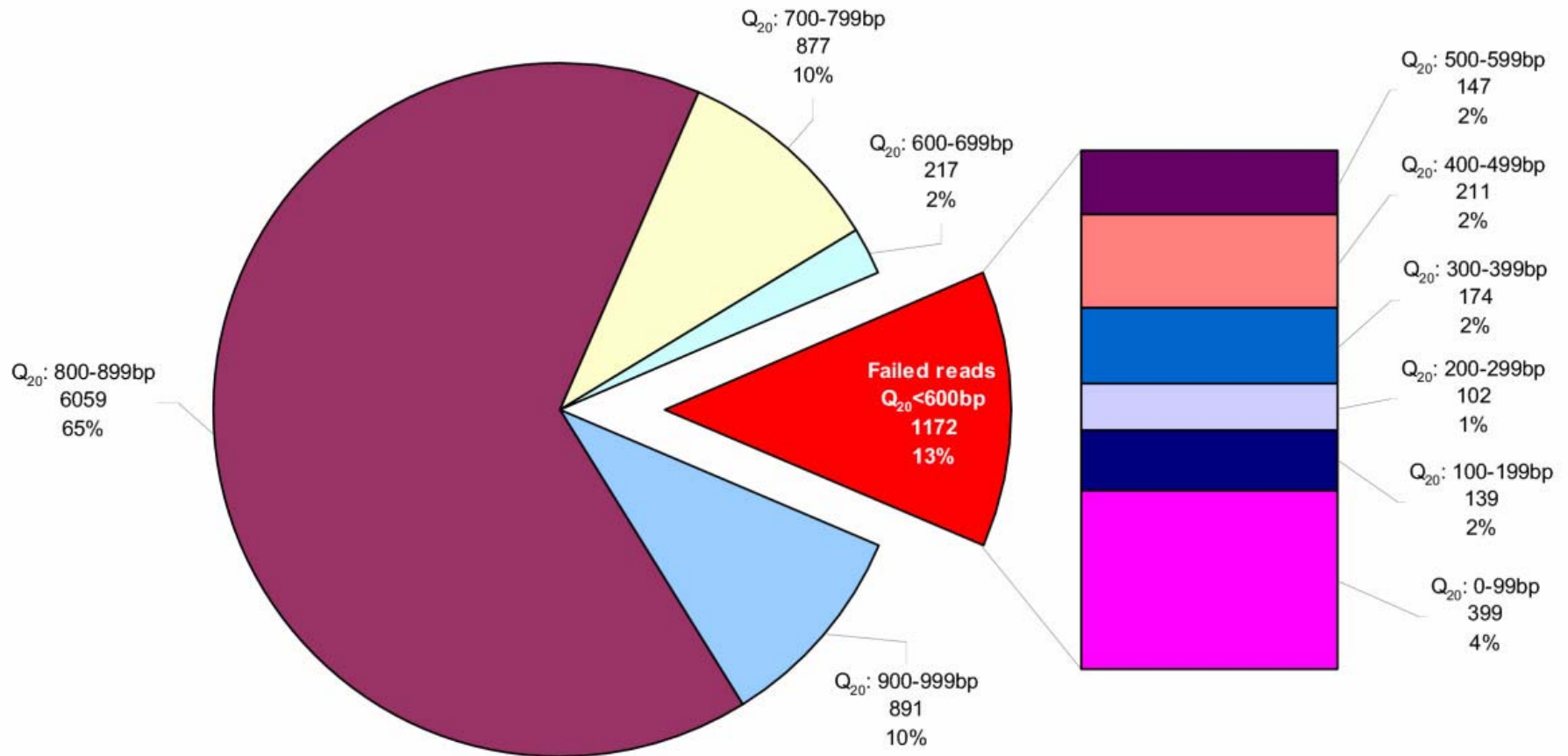
Visual volume verification of DNA prep solutions added

QC gel of DNA template quality

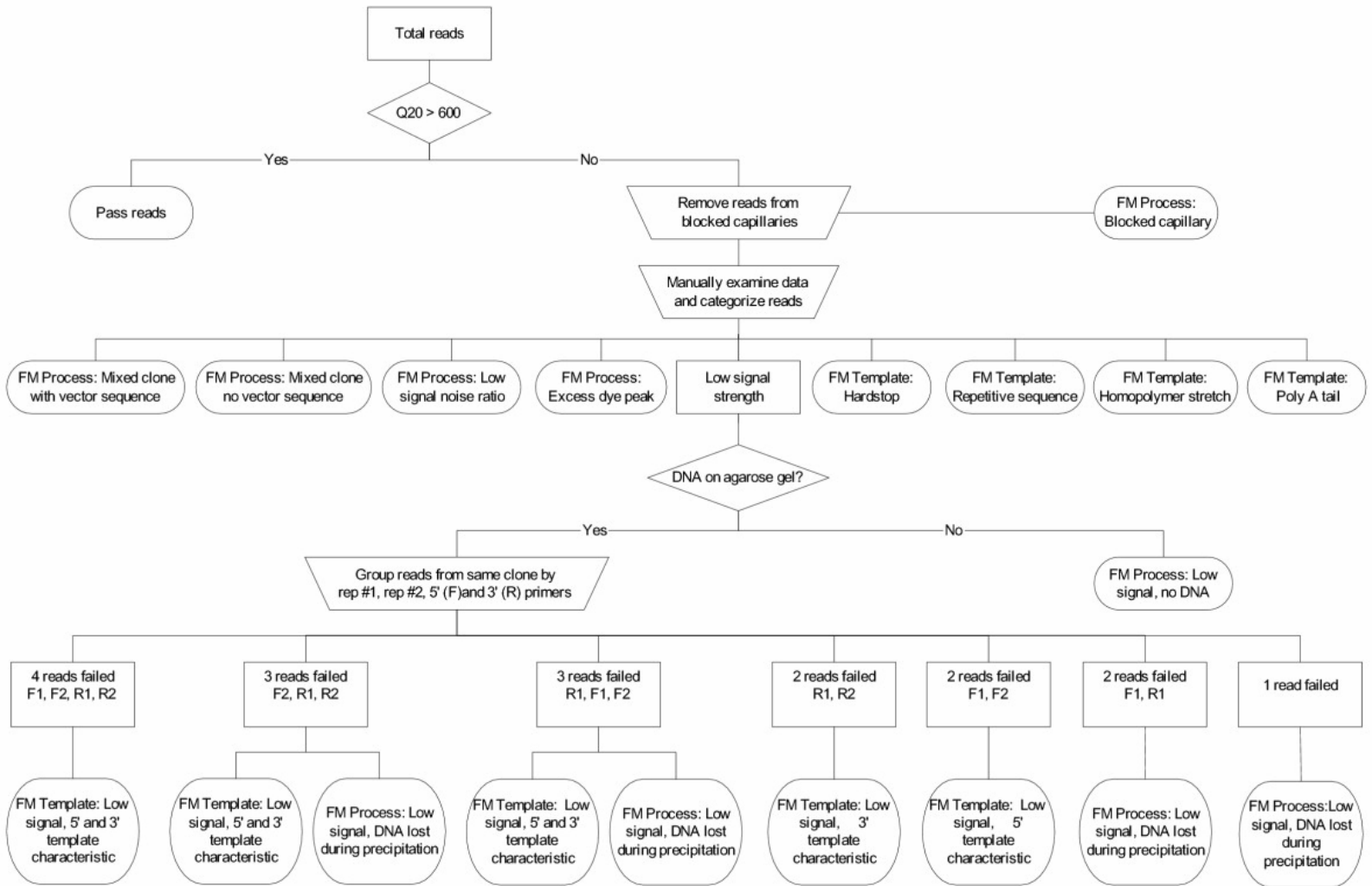
1) Visual volume verification of DNA prep solutions added
 2) Validate precipitation process through process of elimination

Initial run on all sequencers with in-house standards to check for plugged capillaries

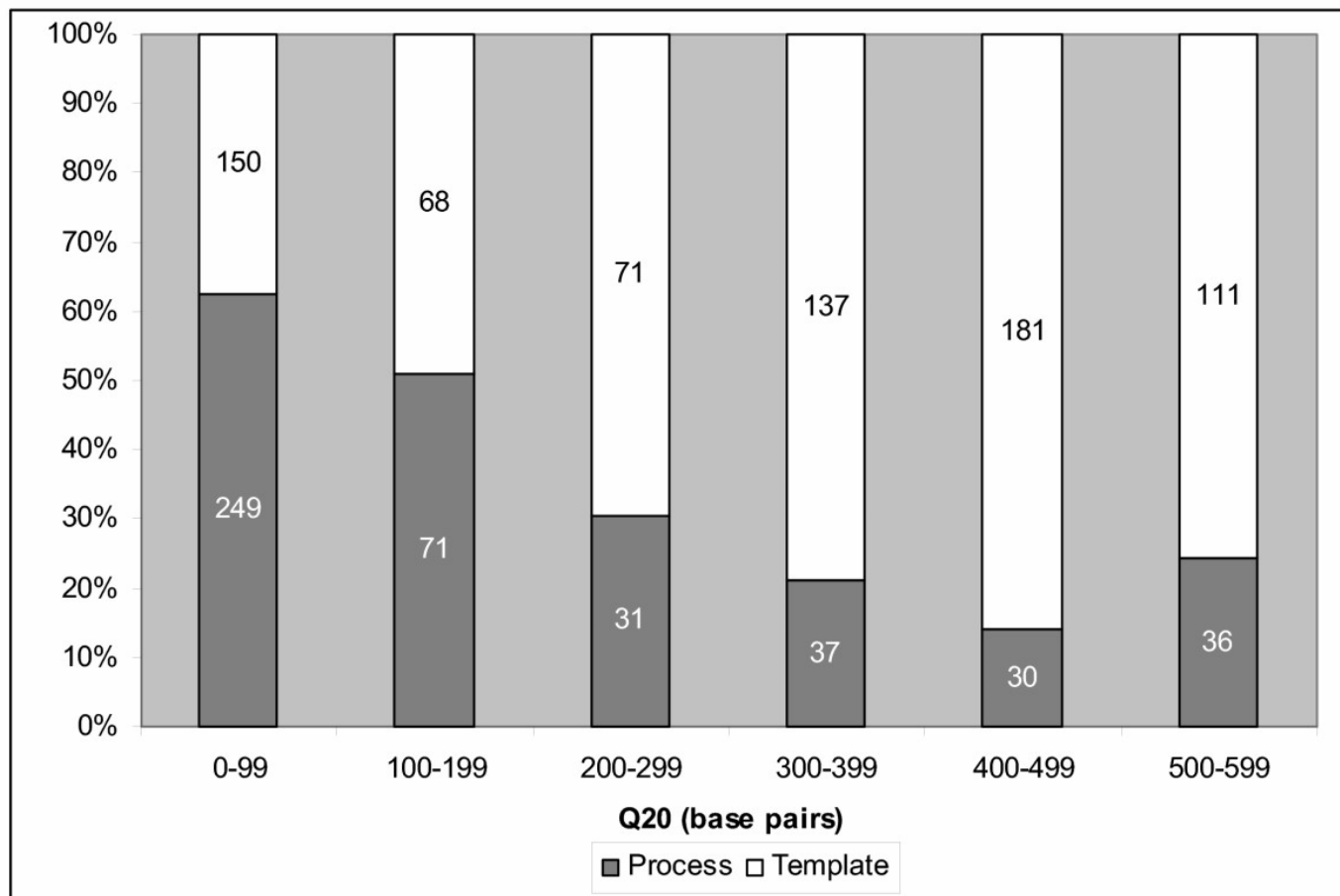
Average read length breakdown



Analysis pipeline

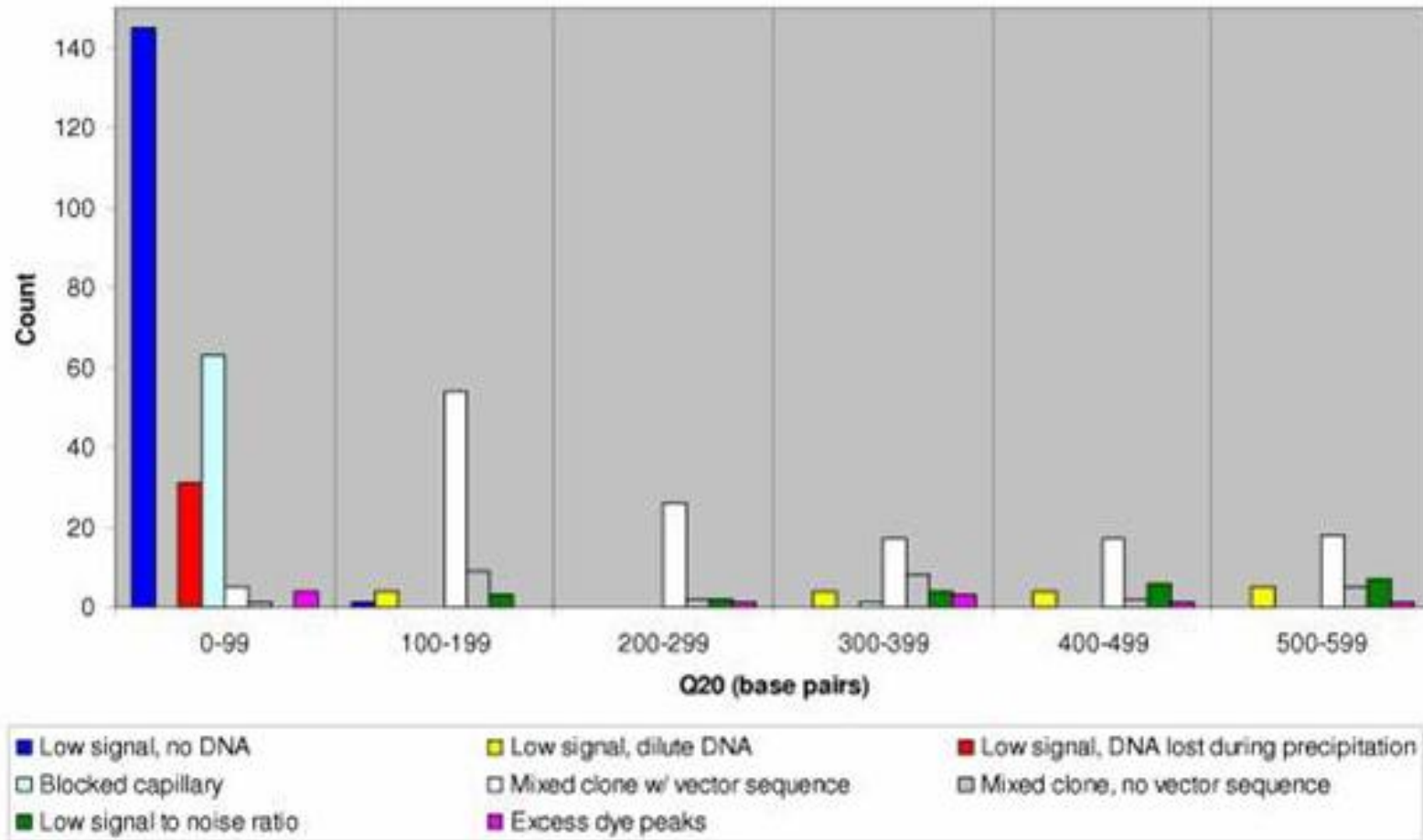


Distribution of process vs. template failed reads

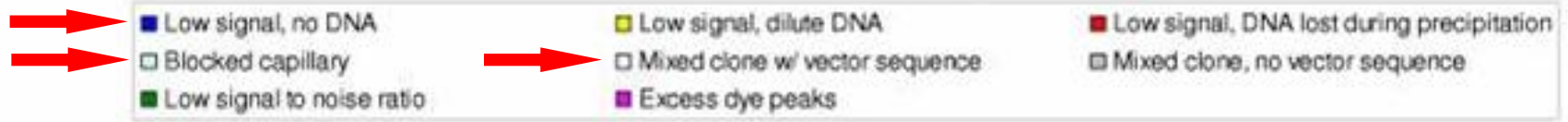
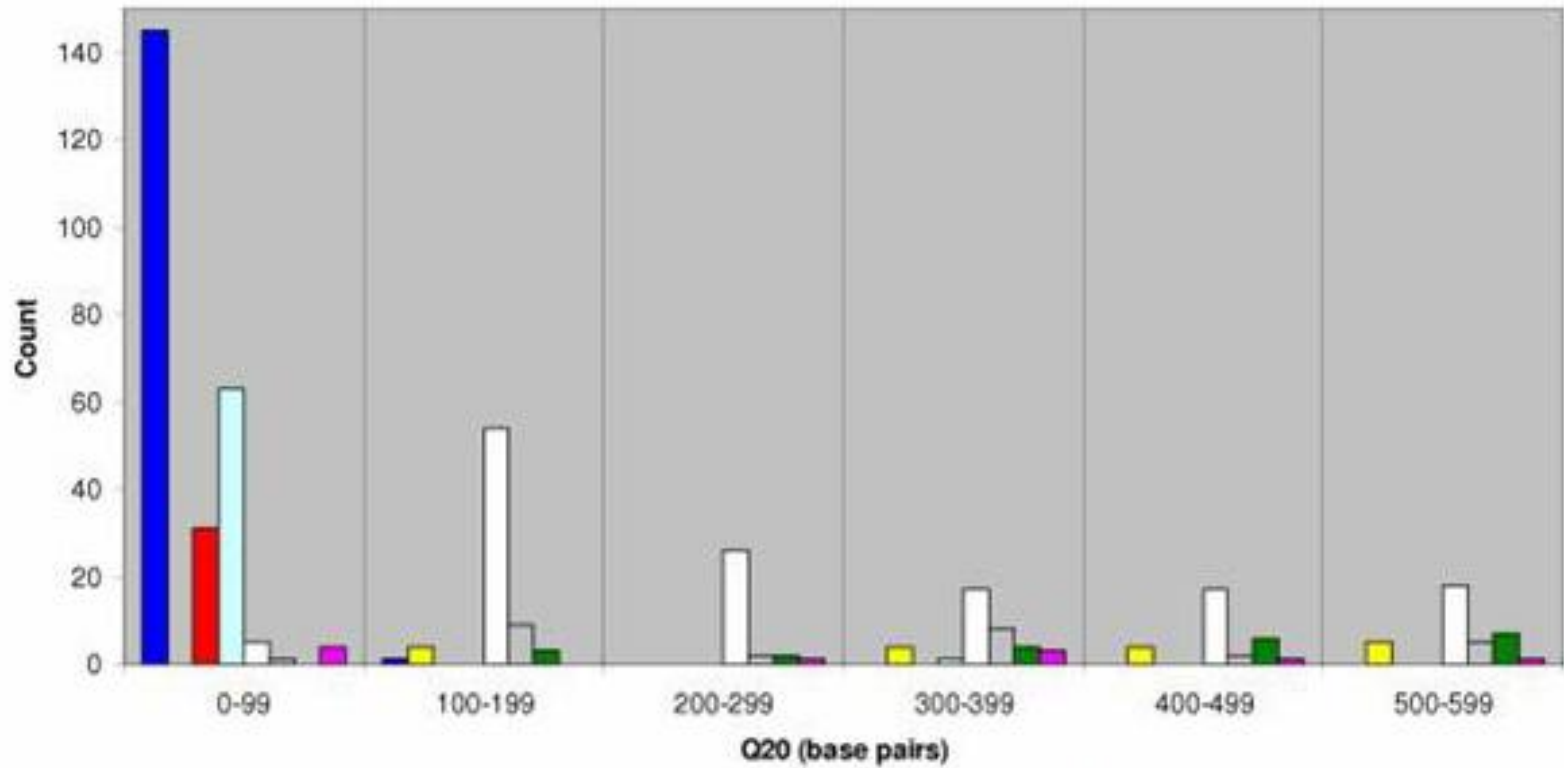


- 454/1172 (39%) of total failed reads were due to process-related failures. Represents 4.9% of total 9,216 overall reads.

Process related failure mode distribution



Process related failure mode distribution



Key pipeline modifications

- **Low signal, no DNA** (146 reads, 12.5% of total failed reads)
 - > Adjustments made picking parameters of QPIX colony picker
 - > DNA preparation adjustments, limited to adjusting dispense volumes
- **Mixed clone with resolvable vector insert** (137 reads, 11.7% of total failed reads)
 - > Increase wash cycles between transfers during QPIX colony picking process and Biomek FX DNA transfers during reaction setup
- **Blocked capillaries** (64 reads, 5.5% of total failed reads)
 - > Shorten time between array swaps

Future direction

- Use other templates test sets e.g. SAGE and shotgun libraries. Compare results obtained from different libraries
- Re-evaluation of our new automated sequencing pipeline, based on 400nL total sequencing reaction volume

> Duane E. Smailus, Andre Marziali, Philip Dextras, Marco A. Marra, and Robert A. Holt. **Simple, robust methods for high-throughput nanoliter-scale DNA sequencing.** *Genome Res.*, Oct 2005; 15: 1447 - 1450.

Acknowledgements

Sequencing Production Group:

- Ryan Babaikaff
- Mabel Brown John
- Ruth Featherstone
- Josh Morin
- Teika Olson
- David Wong
- Miruna Balasundaram (QA Assistant Coordinator)
- Sarah Barber (Assistant Coordinator)
- Jeff Stott (Coordinator)

Duane Smailus (Tech Development Coordinator)

Marco Marra (Director)

Robert Holt (Head, Sequencing)