

Additional File 1

A novel approach to sequence validating protein expression clones with automated decision making

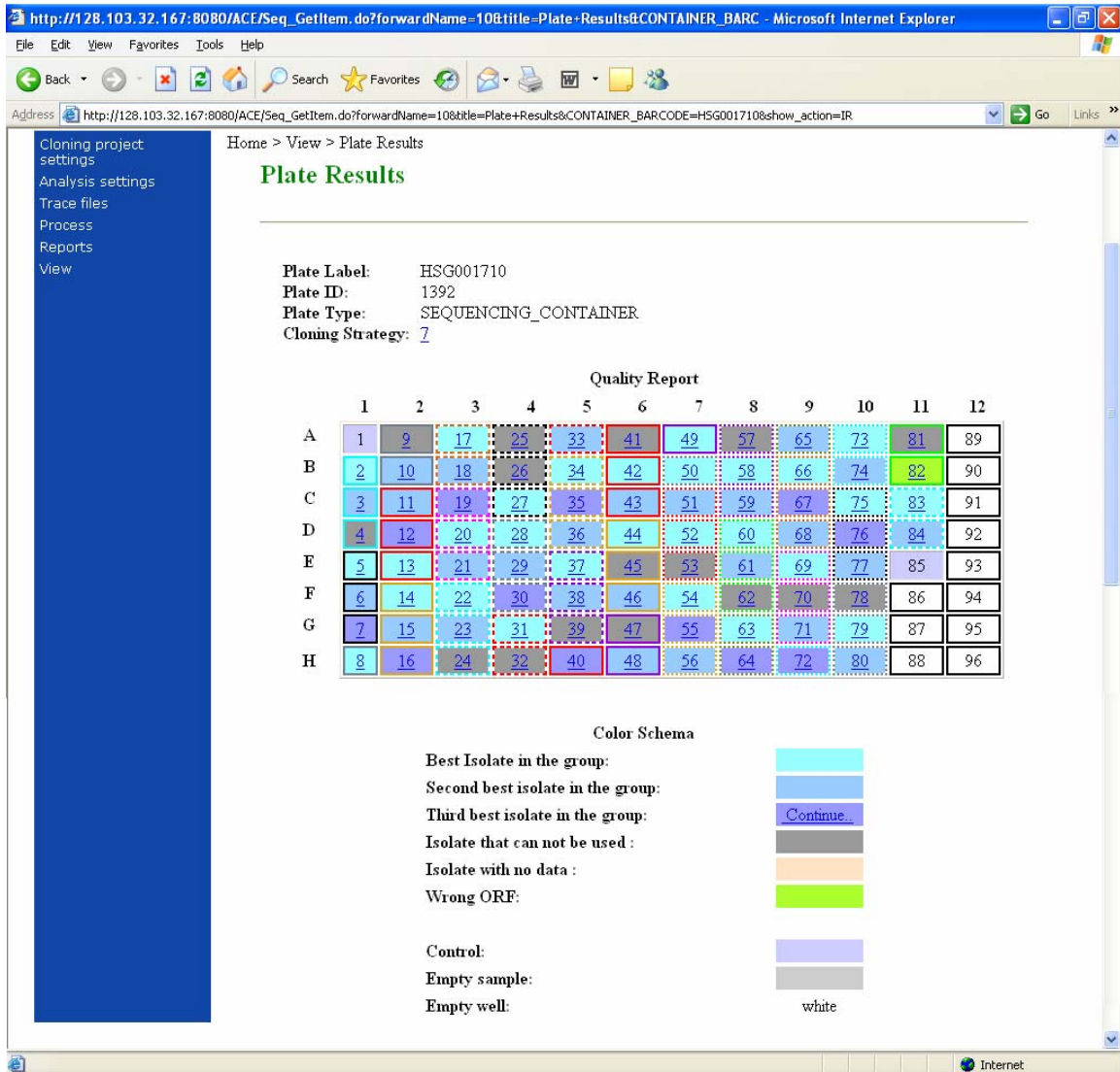
Elena Taycher¹, Andreas Rolfs¹, Yanhui Hu¹, Dongmei Zuo^{1,2}, Stephanie E. Mohr²,
Janice Williamson¹, Joshua LaBaer^{1,2§}

¹ Harvard Institute of Proteomics, Harvard Medical School, 320 Charles St., Cambridge, MA 02141

² DF/HCC DNA Resource Core, Harvard Medical School, 320 Charles St., Cambridge, MA 02141

Contents

Additional File 1, Figure 1. Isolate Ranking Report	2
Additional File 1, Figure 2. Request for Approval of Specific Primers	3
Additional File 1, Figure 3. Decision Tool Execution.....	4
Additional File 1, Figure 4. Create New Set of Parameters for Clone Ranking.....	5
Additional File 1, Figure 5. Online example of Gap Mapper Result.....	6
Additional File 1, Figure 6. Online example of Low Confidence Region Finder Results..	7
Additional File 1, Figure 7. Parameter Settings for Sequencing Primer Design	8



Additional File 1, Figure 1. Isolate Ranking Report.

Screenshot of ACE showing plate results of Isolate Ranker with color-coded clone rank. Identical border style (color and line style) is used to show wells that represent multiple isolates (clones) of the same gene. The color code is indicated below the plate map. As shown here, all isolates for the gene can be placed together, but it is not required. Clicking on any well number will open up a window with detailed information about the clone including the end reads, clone sequence and their alignments with the expected reference sequence.

The screenshot shows a web browser window with the title "<bean:message key='bec.name'/'> : title - Microsoft Internet Explorer". The address bar shows "http://128.103.32.167:8080/ACE/RunProcess.do". The page content is titled "Request for Approve Internal Primers".

On the left side, there is a navigation menu with the following items: settings, Analysis settings, Trace files, Process, Reports, and View. The "Process" menu is expanded, showing sub-items: Upload plate information, Read manipulation, Evaluate clones, Internal primer design and order, Upload clones sequences, Set final clone status, Delete data, and View process results.

The main content area displays three sections for different clones, each with a "Change Status" checkbox, "Primer3 Specification" (a number), and "Reference Sequence Id" (a number). Each section contains a table of primers.

Clone 119382:

- Change Status:
- Primer3 Specification: 87
- Reference Sequence Id: 42746

	Name	Sequence	Tm	Position	Orientation	Status	Submission Type	Submitter
<input type="checkbox"/>	F1	GCTCAAGACCATGACCGAT	58.603	264	Forward	Designed	Gene specific, calculated	System
<input type="checkbox"/>	F2	ATGCGATGCTCTCTCATCA	57.939	622	Forward	Designed	Gene specific, calculated	System

Clone 88:

- Change Status:
- Primer3 Specification: 88
- Reference Sequence Id: 42746

	Name	Sequence	Tm	Position	Orientation	Status	Submission Type	Submitter
<input type="checkbox"/>	F1	ATCCTCTTCCTCTGACCC	58.028	313	Forward	Designed	Gene specific, calculated	System
<input type="checkbox"/>	F2	GGCCTTTATCACCATCCAG	57.947	654	Forward	Designed	Gene specific, calculated	System

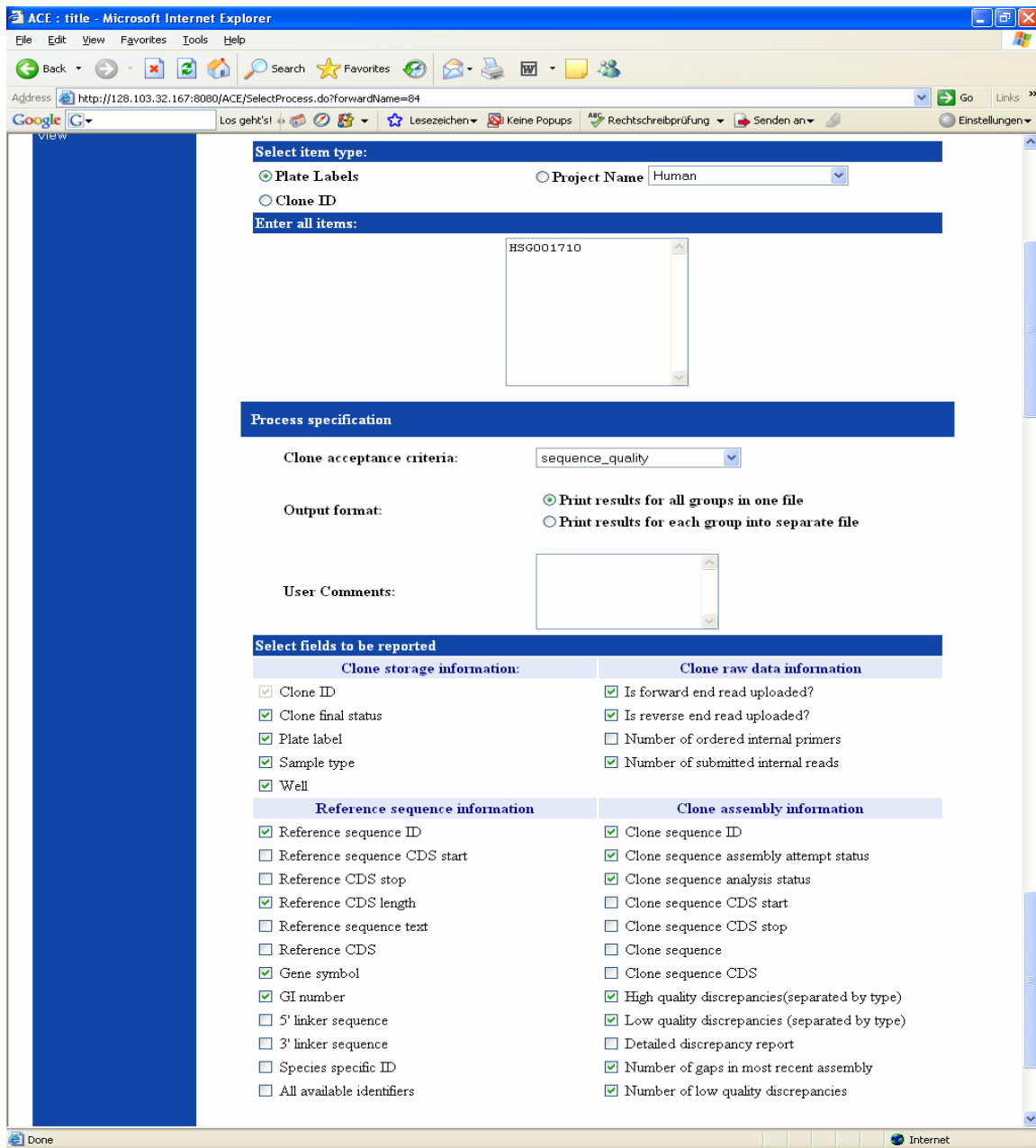
Clone 26:

- Change Status:
- Primer3 Specification: 26
- Reference Sequence Id: 42746

	Name	Sequence	Tm	Position	Orientation	Status	Submission Type	Submitter
<input checked="" type="checkbox"/>	F1	CTCCATCATTGTTTCGTGG	58.973	195	Forward	Approved	Gene specific, calculated	System

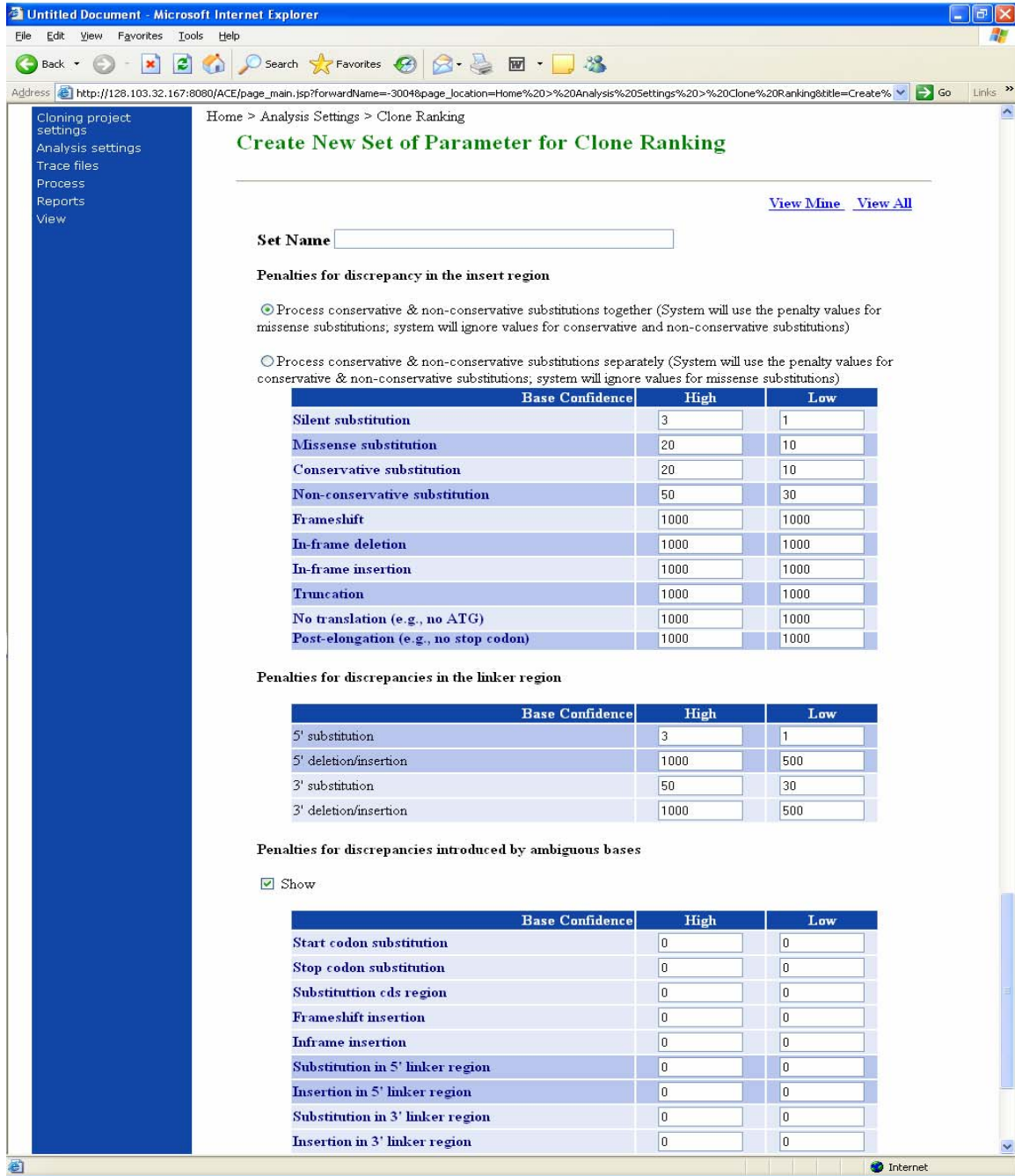
Additional File 1, Figure 2. Request for Approval of Specific Primers.

Screenshot shows the interface for selecting primers for the vendor order. ACE allows the user to design several sets of gene-specific primers for the clone using different Primer Designer specification. Primers from different sets can be put on the same vendor order. Upon placing the order ACE selects all primers approved for the specified clones and puts them in 96-well plate format. The user has an option to specify range of wells in which primers should be placed (preserving empty wells for controls if desired). ACE provides instructions for re-arranging the clone templates so that they match the vendor order and tracks all primers ordered for a particular clone.



Additional File 1, Figure 3. Decision Tool Execution.

Screenshot of user interface that initiates execution of Decision Tool which sorts clones submitted for the analysis into appropriate categories (details in the manuscript). User can define a set of clones for analysis by providing: (1) plate labels; (2) unique clone ID(s); (3) specifying a project. Users invoke predefined acceptance criteria for the analysis by indicating the name of the saved set by selecting it from a pull down list of available sets (see text and Supplemental Figure 3). The output can be directed to separate files (useful for managing groups of clones for downstream steps) or to a single file (to obtain an overall project summary). Users can also specify which information they would like to include in the Decision Tool Report (e.g., clone sequence, clone CDS, discrepancy descriptions). An example of a decision tool report is included in the supplementary material.



Additional File 1, Figure 4. Create New Set of Parameters for Clone Ranking. Screenshot of interface to create a set of parameters for isolate ranking. This is particularly relevant for workflows in which multiple isolates for each gene are selected with the intent to pick one “best candidate”. For each combination <discrepancy type, confidence> (where “confidence” can be “low” or “high”), the user specifies the penalty per discrepancy combination.

Cloning project settings
Analysis settings
Trace files
Process
Reports
View

Home > Process > View Results > View Latest Contig Collection

View Latest Contig Collection

Clone ID: 120016
Reference Sequence ID: 42786

Contig Name	Contig Type	Contig ID	CDS Start	CDS Stop	LQR defined	Alignment	Discrepancy Report
Gap 1	Gap	15467	-1000	844			
Contig 1	Contig	15465	845	1154	Yes	<input type="button" value="Alignment"/>	<input type="button" value="Discrepancy Report"/>
Gap 2	Gap	15468	1155	2462			
Contig 2	Contig	15466	2463	3295	Yes	<input type="button" value="Alignment"/>	No discrepancies
Gap 3	Gap	15469	3296	-1000			

Clone ID: 134915
Reference Sequence ID: 42988

Contig Name	Contig Type	Contig ID	CDS Start	CDS Stop	LQR defined	Alignment	Discrepancy Report
Contig 1	Contig	36250	-49	35	Yes	<input type="button" value="Alignment"/>	No discrepancies
Gap 1	Gap	36252	36	526			

Additional File 1, Figure 5. Online example of Gap Mapper Result.

Screenshot shows contigs and gaps defined for a clone by Gap Mapper. Contigs are characterized by the position on the clone reference CDS and their sequences are analyzed by Discrepancy Finder to allow assessment of clone quality. LQR indicates Low Quality Region.

The screenshot shows a web browser window displaying the Automatic Clone Evaluation (ACE) application. The page title is "Automatic Clone Evaluation (ACE)". The navigation menu includes "About", "Help", "Contact Us", and "Log Out". The breadcrumb trail is "Home > Process > View Results > View Low Confidence Regions for Clone Sequences". The main heading is "View Low Confidence Regions for clone sequences".

On the left, there is a sidebar menu with the following items: Cloning project settings, Analysis settings, Trace files, Process, Reports, and View.

The main content area displays the following information:

- Clone Id: 139940
- Reference sequence ID: 42840
- Clone sequence ID: 55127
- Clone sequence analysis status:
- Clone sequence alignment: [Alignment](#)
- Specification for LQR definition: [32](#)

Below this information is a table with the following columns: Name, CDS Region, Sequence Region, Sequence, and Discrepancy Report.

Name	CDS Region	Sequence Region	Sequence	Discrepancy Report
LQR 1	513 - 745	562 - 794	<pre> GTGGGACCGCTCTACTTCGGCGAGAGGCCAGGGGTAAA GAGGCCTA:TAGGGGGCGGGCCCGCCCGTCTGTACC GGTCTCTGGTCTTGGGGTAAAGACACCTGGGAGAGGGCA GCACCGTCTTACTCGTGGCCACAGCCA:AAATCAAGCCCA GGGGTCTGGGAGAGCGGGGAGGGGCTCTTGGTGGGAC ACCTCCCTCTGGGCGGGCGGGGGGGGGGGGGGGGGGGGG </pre>	Discrepancy Report

Additional File 1, Figure 6. Online example of Low Confidence Region Finder Results.

Screenshot shows example of low confidence regions for clone sequence. Low confidence regions are defined for a sequence by applying a ‘sliding window’ algorithm (see manuscript for details), each region is characterized by position on clone reference CDS. Not viewable because it is at the bottom of the screen is the color code indicating the confidence score.

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

Address http://128.103.32.167:8080/ACE/page_main.jsp?forwardName=-3005&page_location=Home%20>%20Analysis%20Settings%20>%20Primer%20Designer&title=Create* Go Links

Home > Analysis Settings > Primer Designer

Create New Set of Parameter for Primer Design

[View Mine](#) [View All](#) [Delete](#)

Set Name

Primer Design Parameters

Primer Length (bp)	Min: <input type="text" value="18"/>	Optimal: <input type="text" value="21"/>	Max: <input type="text" value="27"/>
Primer Tm (C)	Min: <input type="text" value="57"/>	Optimal: <input type="text" value="60"/>	Max: <input type="text" value="63"/>
Primer GC%	Min: <input type="text" value="30"/>	Optimal: <input type="text" value="50"/>	Max: <input type="text" value="70"/>

Sequencing Parameters

Check here if NO 5p Universal Primer is used (The most upstream forward PCR primer is used when No 5p Universal Primer is checked)

Check here if NO 3p Universal Primer is used

Distance between 5p Universal Primer and START codon (For a left primer, primer start position is the position of the leftmost base) bases

Distance between 3p Universal Primer and STOP codon (For a right primer, primer start position is the position of the rightmost base) bases

Estimated high quality read length (ERL) bases

Window size for testing primers bases

Distance between sequencing primer and start of high quality read length bases

Number of strands to sequence

Single Strand (Coding strand, forward primers only)
 Single Strand (Compliment to coding strand, reverse primers only)
 Both Strands (Both forward and reverse primers until meet in middle)
 Both Strands (Both forward and reverse primers, double coverage)

Done Internet

Additional File 1, Figure 7. Parameter Settings for Sequencing Primer Design.

Screenshot shows interface that allows users to create a set of parameters for primer design which include: (1) sequence related parameters (e.g., Tm, GC content); (2) sequence processing parameters (e.g., length of reliable part of sequencing reads); (3) type of desired coverage (e.g., single forward, single reverse, double coverage).