Documentation for *CopyCount-CNV*™

Software version: 2.1.9.2
Release Date for *CopyCount-CNV*™: November 2, 2016

Written by: John SantaLucia, DNA Software, Inc.
Documentation version: May 5, 2017

CopyCount-CNV™ and DNA Software products are for Research Use Only.

CopyCount-CNV™ Documentation

Step 0: What you need before running *CopyCount*:

A. You will need to retrieve your raw fluorescence PCR data from the Fluidigm BioMark instrument. In the BioMark real time-pcr data analysis software, click the menus:

File>Export and select: Table Results with Raw Data (*.csv)

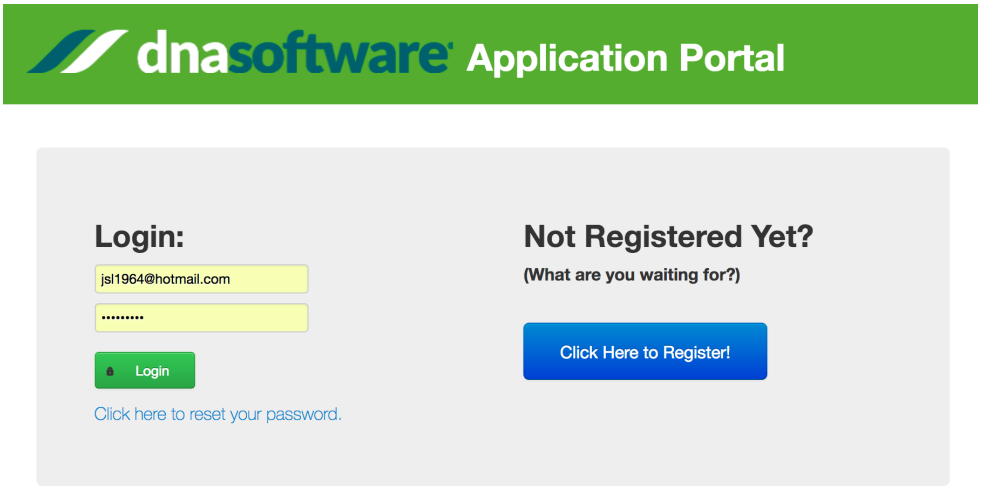
Be sure to save the file with a name that is meaningful and note the location of the file so that you can retrieve it later when running *CopyCount*.

B. To run *CopyCount* you will need to gather the following information about each of the Gene-of-interest and control Assays:

Detection Channel used for Gene-of-interest (typically FAM-MGB)
Detection Channel used for CNV control (typically VIC-MGB)
Reaction Volume: 6.7E-9 (given in Liters)
Length of amplicon: 105 (given in base pairs)
Double Stranded? Yes
If your original target is double-stranded DNA, then select yes (checked box in CopyCount).
Alternatively, select No if the original target is single-stranded DNA or RNA (e.g. cDNA)
Detection Method: TaqMan (can be Dye based (SYBR, EvaGreen, etc.) or TaqMan)
[Primers]: 900E-9 (Molar concentration)
[Probe]: 200E-9 (Molar concentration)
MGB: Yes (Answer "yes" if your TaqMan probe has a minor groove binder at the 3'-end)

Note: When you run an assay once in CopyCount, it is saved, so that in future runs you will not need to look up the assay information. This information is used by *CopyCount* to estimate the absolute calibration if no calibration plate was performed.

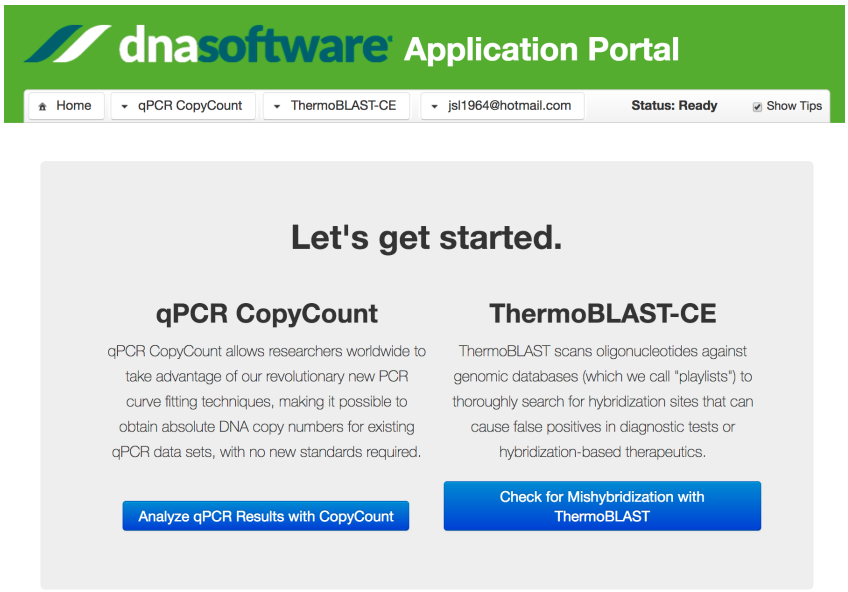
Input Step 1: Go to the DNA Software Application Portal: https://portal.dnasoftware.com/login.jsf



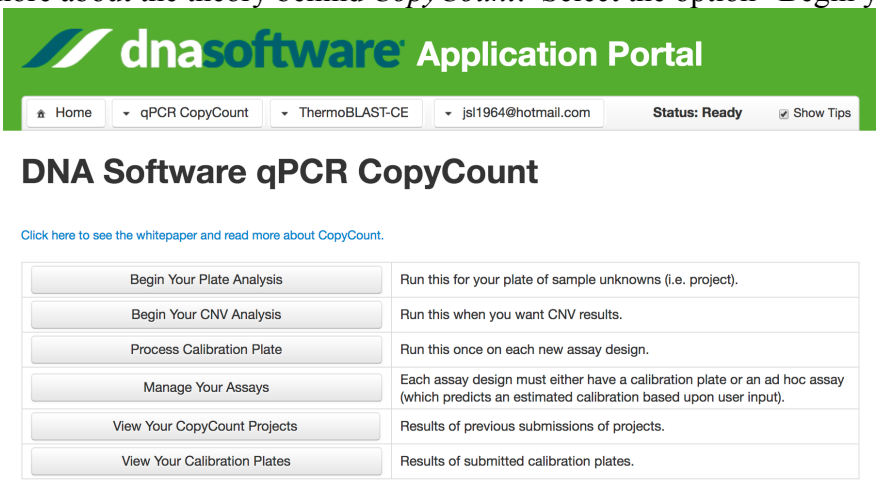
Enter your login Email and Password and hit the Green "Login" button.

If you have never Registered before, it is easy! See the Appendix for Registration Details.

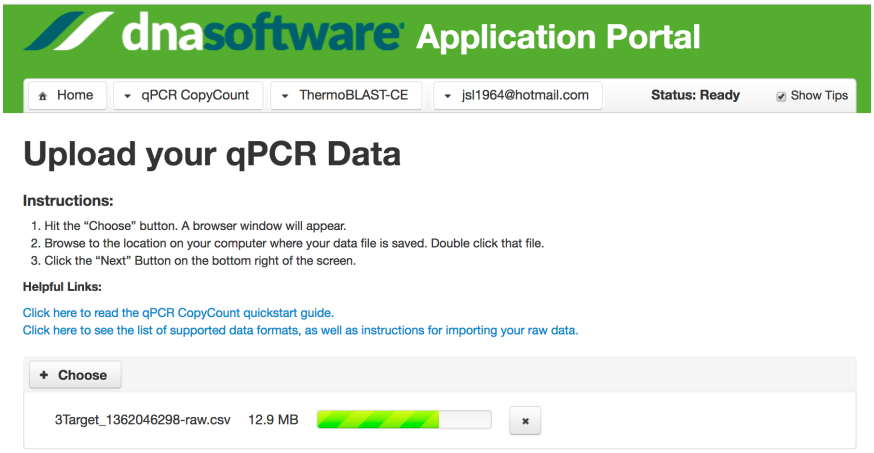
Input Step 2: This will bring you to the "Let's get started" page. Click the Blue button "Analyze qPCR Results with CopyCount".



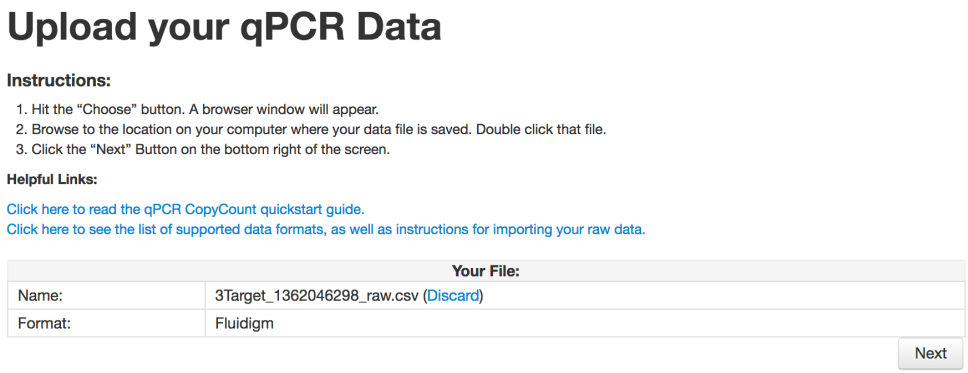
Input Step 3: This will take you to the page below. Note the Blue Link that takes you to a PDF document that describes more about the theory behind *CopyCount*. Select the option "Begin your CNV Analysis".



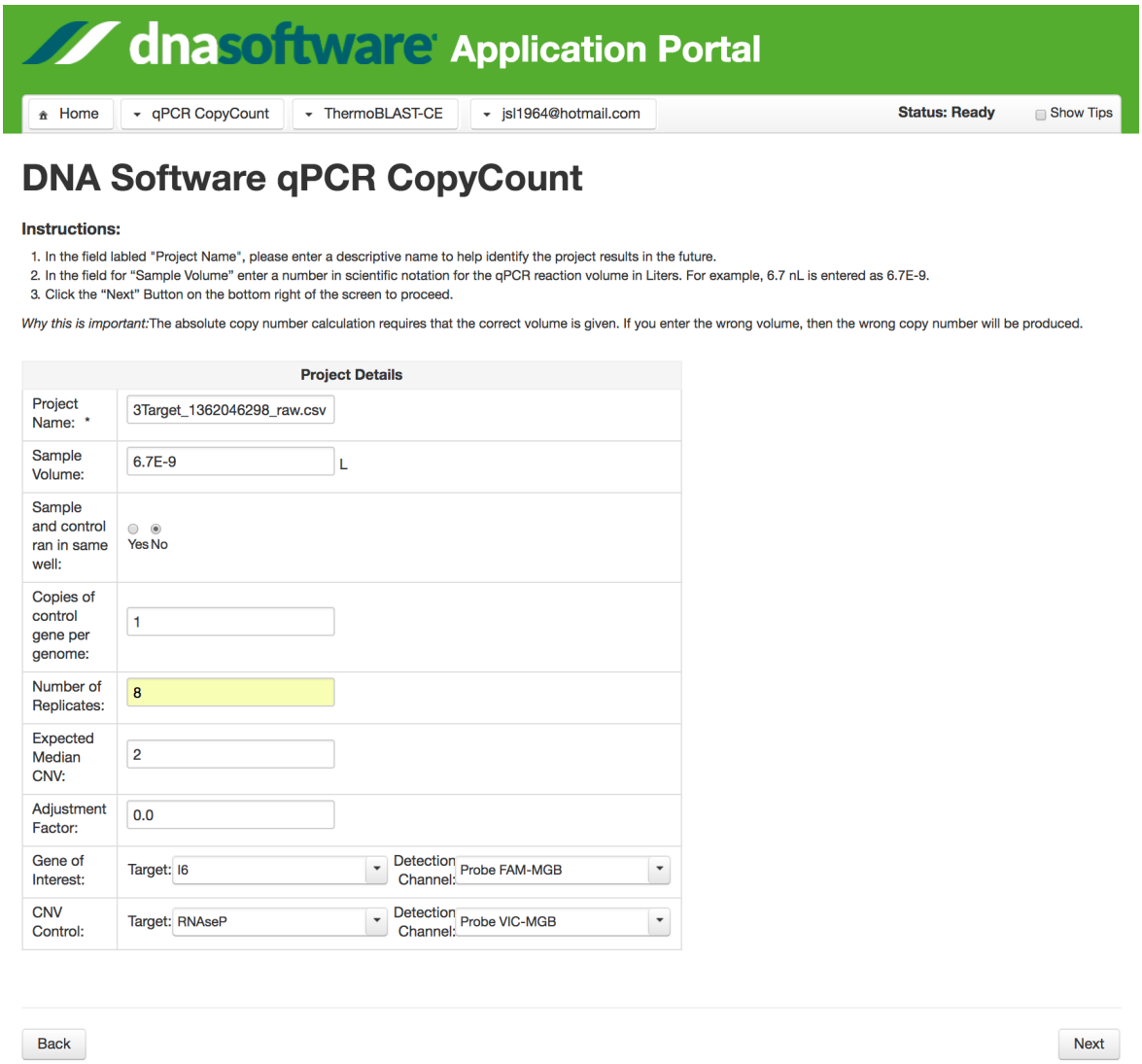
Click on the button "Begin Your CNV Analysis". That will bring you to the "Upload Page". Click on the "+ Choose" button and go to the location where you saved you raw data file (from **Step 0A**).



If your file is recognized by *CopyCount*, then it will give the Format type for the file (in this case it determines it to be "Fluidigm" format). If your file is NOT recognized by *CopyCount*, then that means that you likely uploaded the wrong file (commonly users will mistakenly upload the C_T analysis file rather than the raw PCR fluorescence file). Click on the "Next" button.



Input Step 4: This brings you to the "Project Details" Page (more information on next page). Be sure to fill in all the fields (i.e. use information from **Step 0B**):



If you are a new user, then you can select the "Show Tips" box at the top right hand side of any page. We recommend unchecking "Show Tips" as the default and click it only when you have explicit questions.

Details about inputs:

Project Name: I6 3Target_1362046298
CopyCount automatically fills in the name of the raw data file (e.g. 3Target_1362046298_raw.csv). But we recommend that you put in an informative name, since the same raw data file can be used for multiple assays that are processed separately. In this case, I named it "I6 3Target" because this is for samples using the I6 assay from plate number 1362046298.

Sample Volume: 6.7E-9 (give the PCR reaction volume in Liters, use scientific notation)

Sample and Control Run in the Same Well: Yes xNo (Default = No)
The user can set up their CNV experiment to run the gene-of-interest and the CNV control in either the same well (with different fluorophores) or in separate wells. *CopyCount* performs different replicate averaging and statistics for these two cases (ratio of the averages vs. average of the ratios are different!). Currently, Fluidigm does not support running the GOI and control in the same well (since it requires special instrument calibration with color compensation), but that capability is under

development. Acquiring CNV GOI and control data in the same well is the preferred best practice since it results in 41% smaller CNV error bars.

Number of Replicates: 8 (integer)
CopyCount can accept any integer up to the number actually acquired. In this case, we acquired 8 replicates for the GOI and the Control. If you select a number less than the actual number acquired, then *CopyCount* will break up the actual replicate set into smaller replicate sets. For example, if the actual number of replicates collected was 8, and you set the parameter Number of Replicates to "2", then *CopyCount* will break up the data into 4 groups of 2 replicates (each of the 4 groups is given a different name by *CopyCount*).

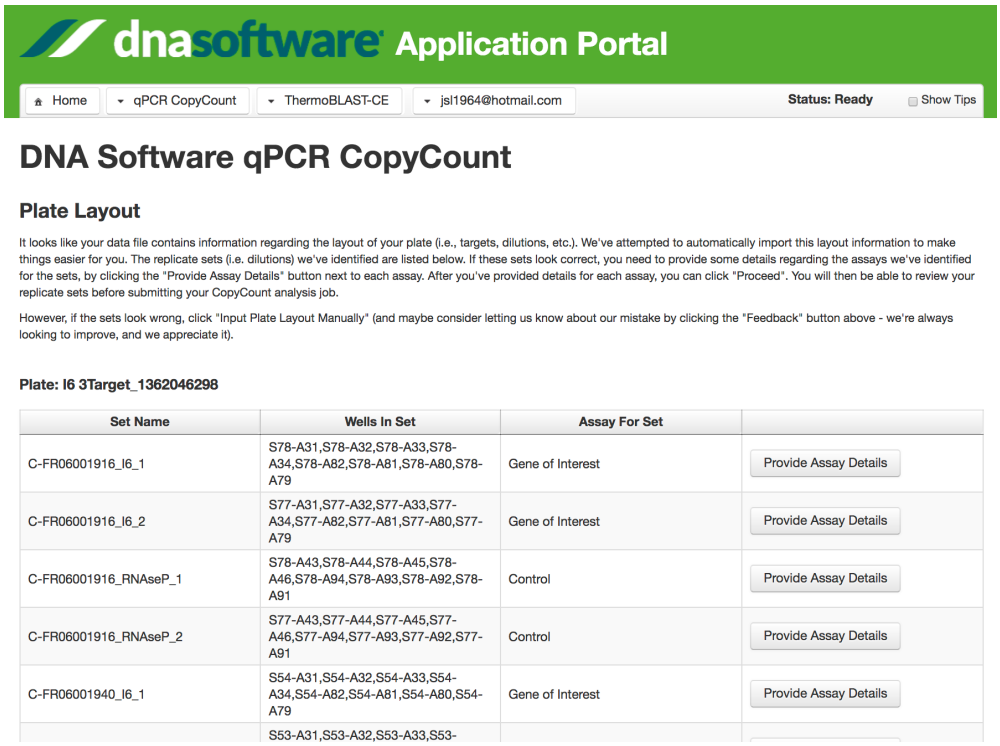
Expected Median CNV: 2 (integer)
Most users will just leave the default value of "2". Choose a number here that represents the number that you think will be most often represented CNV value in your data set. If you think most of the wells have a CNV = 2 (i.e. homozygote), then set this parameter to "2". Note that this is used just to get the algorithm started for the "Adjustment Factor" (described below). If you enter a slightly wrong value like "1" or "3" the algorithm will still compute correctly. However, if you are using *CopyCount* to compute CNV values larger than 3, then it is important to give a value that is close to the actual most common CNV value.

Adjustment Factor: 0.0 (real number, default = 0.0)
Most users will just leave the default value of 0.0. The "Adjustment Factor" is the factor used to convert the raw ratio of Copies of GOI / Copies of Control to compute the correct integer CNV value. If this is set to 0.0 then the algorithm will automatically determine the Adjustment factor (that value is in the CNV output file). If the user provides a value (like from a previous run of *CopyCount*) then that value will be used as a fixed Adjustment factor. In the (very rare) event that *CopyCount* gives an Adjustment factor that is "wacky", then this give the user a mechanism to force a value.

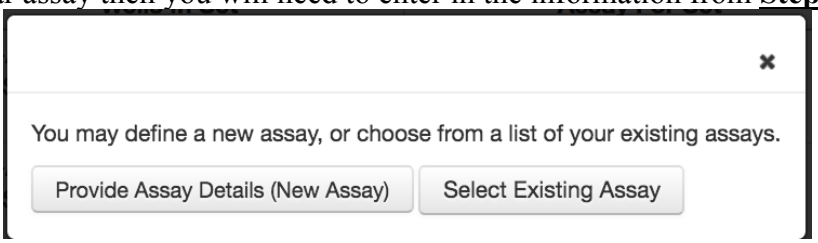
Gene of Interest: Target: I6 Detection Channel: Probe FAM-MGB
CNV Control: Target: RNaseP Detection Channel: Probe VIC-MGB
Use the menu pull downs to select all 4 of the required inputs.
It is VERY important to get this correct!! Wrong selections will fit the wrong data!

Click on "**Next**" Button.

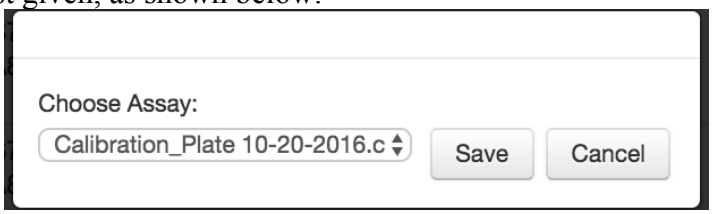
Input Step 5: *CopyCount* then automatically detects all the sample replicate sets. This is shown on the "Plate Layout" Page:



You cannot go forward until you Click on the "Provide Assay Details" once for the Gene of Interest, and once for the Control. Click on one of the buttons (it doesn't matter which one). If this is the first time you are running a particular assay then you will need to enter in the information from **Step 0B**.



If you have previously made an Assay for your GOI, then you can select the "Select Existing Assay" option and just select from the list given, as shown below:



If you select "Provide Assay Details (New Assay), then you get the following window:

1. Enter a name for the new assay.
2. Enter the amplicon length (typically 60 to 150 BP, but longer lengths are allowed).
3. If your target DNA is double-stranded, then check the associated box.
4. Choose whether your assay is a TaqMan assay or DNA-binding dye (e.g. SYBR or SYTO). Other qPCR formats (e.g. Molecular beacons and Scorpions) are not yet supported.
5. Enter the primer concentration (molar units). If your assay is a TaqMan assay, then you will also need to enter the probe concentration.
6. If your TaqMan probe has a minor groove binder (e.g. for assays from Life Technologies), then check the box for MGB.

Assay Details

Name: *	I6 new
Amplicon Length: *	71
Double-stranded?	✔
Detection Method:	TaqMAN
Primer Concentration	900E-9 M
Probe Concentration	200E-9 M
Using MGB?	✔

Cancel Save

After you select the "Save" button, then you will be taken back to the Plate Layout Page. Note that all the replicates for the GOI are now filled in with "I6 new".

DNA Software qPCR CopyCount

Plate Layout

It looks like your data file contains information regarding the layout of your plate (i.e., targets, dilutions, etc.). We've attempted to automatically import this layout information to make things easier for you. The replicate sets (i.e. dilutions) we've identified are listed below. If these sets look correct, you need to provide some details regarding the assays we've identified for the sets, by clicking the "Provide Assay Details" button next to each assay. After you've provided details for each assay, you can click "Proceed". You will then be able to review your replicate sets before submitting your CopyCount analysis job.

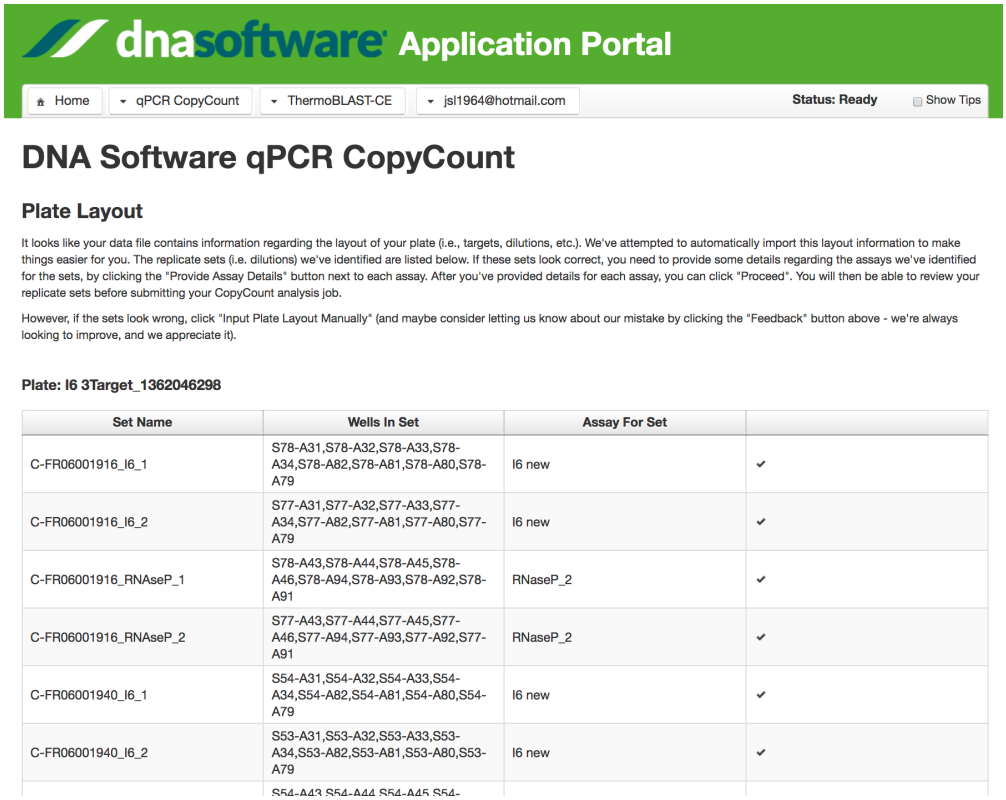
However, if the sets look wrong, click "Input Plate Layout Manually" (and maybe consider letting us know about our mistake by clicking the "Feedback" button above - we're always looking to improve, and we appreciate it).

Plate: I6 3Target_1362046298

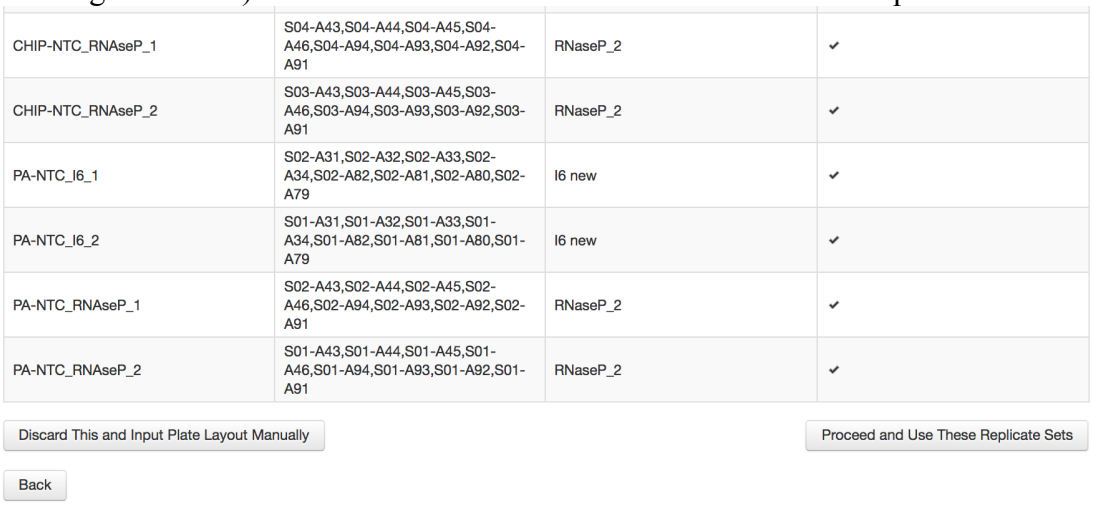
Set Name	Wells In Set	Assay For Set	
C-FR06001916_I6_1	S78-A31,S78-A32,S78-A33,S78-A34,S78-A82,S78-A81,S78-A80,S78-A79	I6 new	✔
C-FR06001916_I6_2	S77-A31,S77-A32,S77-A33,S77-A34,S77-A82,S77-A81,S77-A80,S77-A79	I6 new	✔
C-FR06001916_RNAseP_1	S78-A43,S78-A44,S78-A45,S78-A46,S78-A94,S78-A93,S78-A92,S78-A91	Control	Provide Assay Details
C-FR06001916_RNAseP_2	S77-A43,S77-A44,S77-A45,S77-A46,S77-A94,S77-A93,S77-A92,S77-A91	Control	Provide Assay Details
C-FR06001940_I6_1	S54-A31,S54-A32,S54-A33,S54-A34,S54-A82,S54-A81,S54-A80,S54-A79	I6 new	✔
C-FR06001940_I6_2	S53-A31,S53-A32,S53-A33,S53-A34,S53-A82,S53-A81,S53-A80,S53-A79	I6 new	✔
C-FR06001940_RNAseP_1	S54-A43,S54-A44,S54-A45,S54-A46,S54-A94,S54-A93,S54-A92,S54-A91	Control	Provide Assay Details

However, the Control replicate sets still have the "Provide Assay Details" buttons. Click on one of those buttons. I selected an Existing Assay for "RNaseP_2".

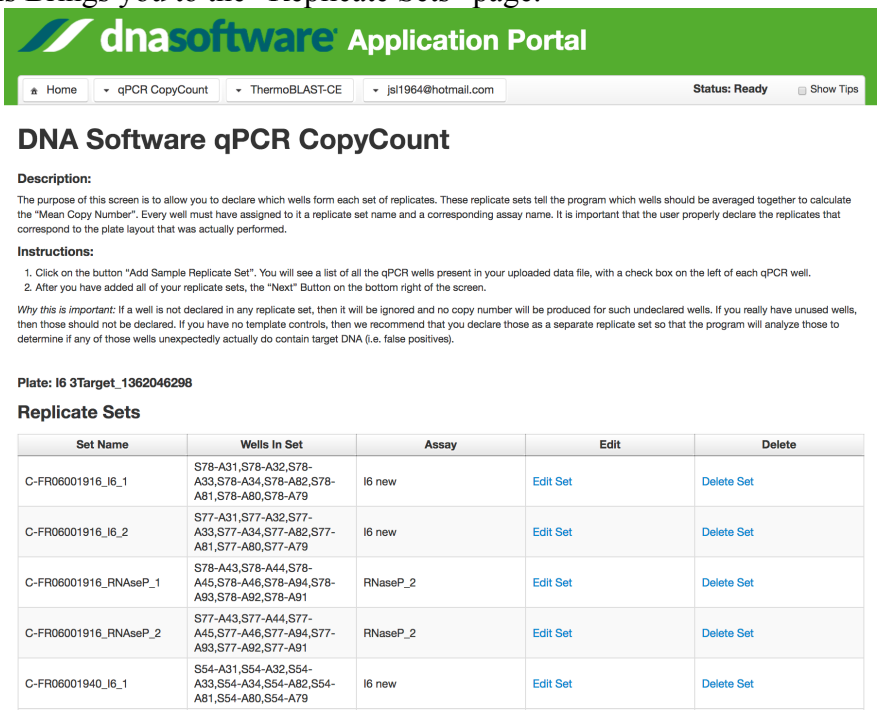
Then the program returns you to the Plate Layout Screen as shown below:



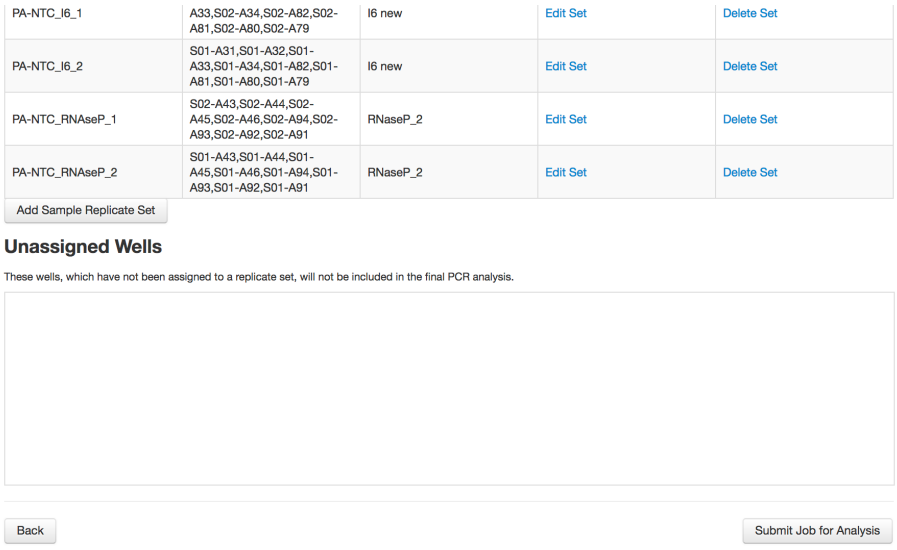
You can see that now all the replicate sets have an Assay Name. In your browser, Scroll to the bottom of the page (this is a huge data file!). Click on the Button "Proceed and Use These Replicate Sets".



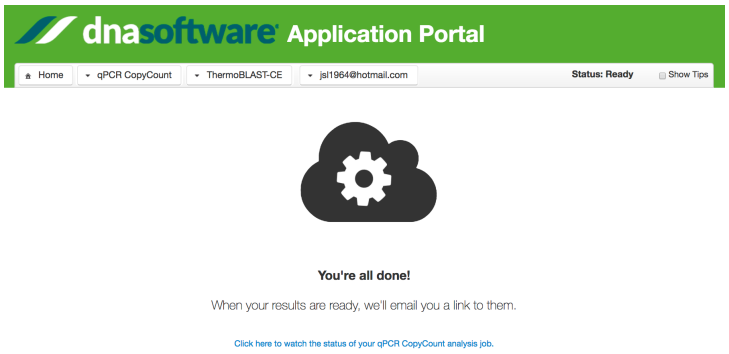
Input Step 6: This Brings you to the "Replicate Sets" page.



Scroll to the bottom of that huge page and select the button "Submit Job for Analysis".

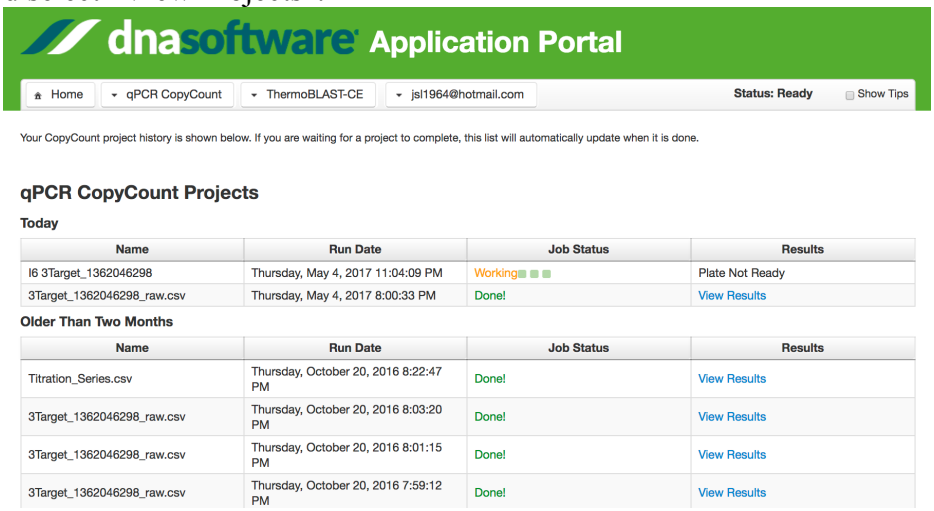


You will get the following screen if you did everything correctly:



Click on the Blue link, to see your results.

Output Step 1: Go to the Results page. (You can get to the anytime by clicking on the menu "qPCR CopyCount" and select "View Projects".



For large Fluidigm datasets the analysis will take about 1-2 minutes at which time the web page will update the Job Status to "Done!".

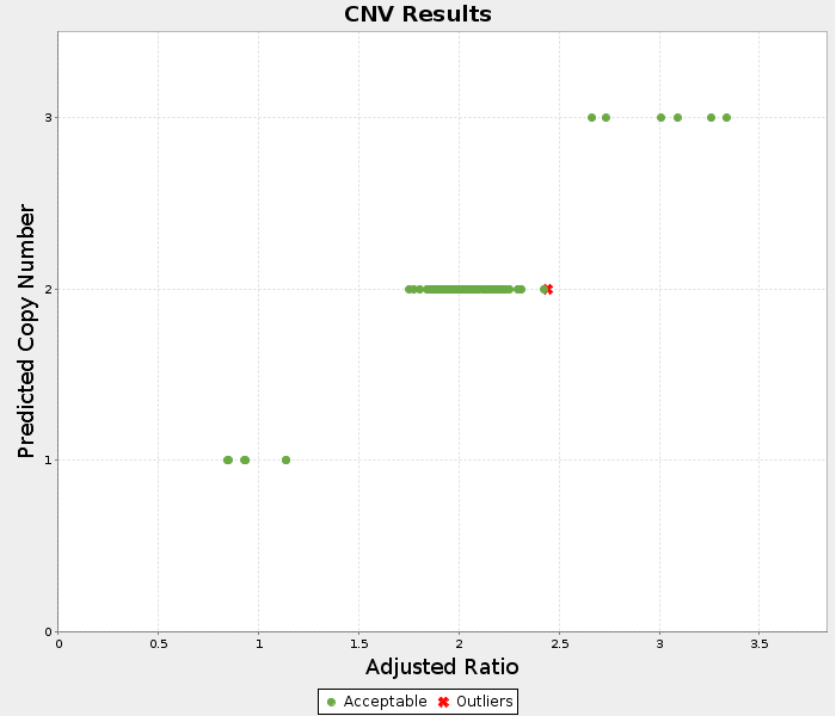
Output Step 2: Click on "View Results" to get the following results page. You can use your browser to scroll down to see more information.

Note that you can download all the tables of data as .CSV files and graphs as .png files (Click on the Blue links). Other information reported are as follows:

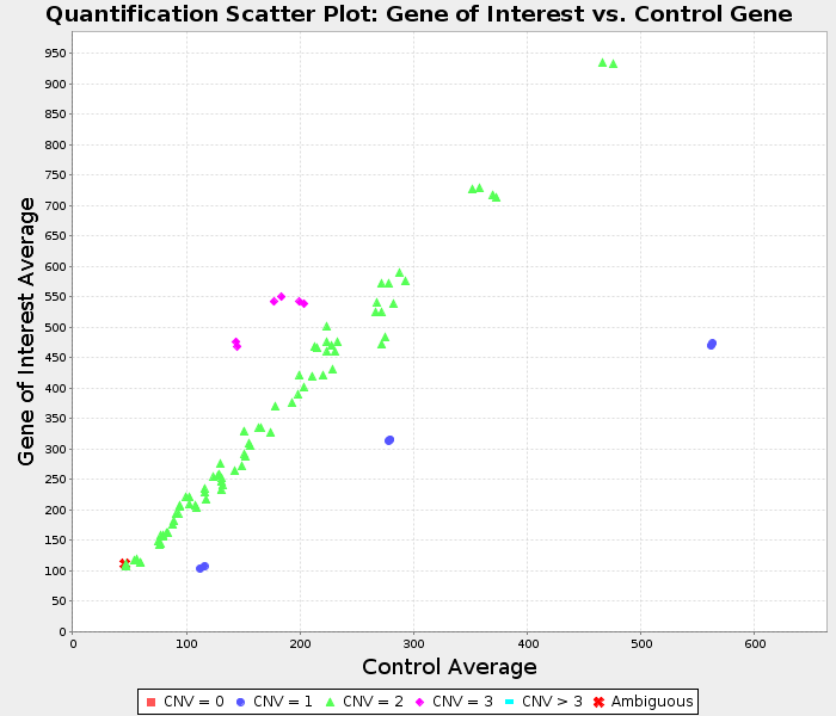
Adjustment Factor = 0.0655 (you could use this number for other data sets in future runs of CopyCount for other samples of I6. This value should not change substantially from plate to plate for the same assay, but different adjustment factors are typical for different assays. Note: we define an "Assay" by the primer sequences.

The Standard Deviation = 0.0832 (this is the standard error in the experimental Ratio of GOI/Control averaged over all the replicate sets). Thus in this case, a CNV of 2.0 is different from CNV =3.0 by 12 standard deviations (Z-score = (3 -2)/0.0832 = 12.02).

The graph below, gives a summary of all of the Predicted CNV (Y-axis) vs. the experimental CNV ratio (X-axis) so you can see the quality of the CNV calls. Note that highly confident CNV calls are shown as green dots "•" and ambiguous CNV calls are graphed with a red "**x**" (CNV values close to half-integers are ambiguous).



The graph below summarizes the scatter in the Copies of GOI (Y-axis) vs. Copies of Control (X-axis). This shows you the spread CNV data. Samples with CNV = 1 lie along the line y=x. Samples with CNV = 2 lie along the line y = 2x. Note that PCR data acquired with very low amount of sample DNA (below 100 molecules) are not as reliable as datasets collected with higher DNA amounts.



An example of the Table of CNV results is shown below:

Number	Replicate Set Names	GOI Average	Control Average	Raw Ratio	Adjusted Ratio	CV of Ratio	Z-Score	Confidence	CNV Call	Comments
1	C-FR06001916_I6 C-FR06001916_RI	45	358	0.127	1.922	0.0865	0.933	0.993	2	Unreliable due to low sample
2	C-FR06001916_I6 C-FR06001916_RI	44	340	0.13	1.968	0.0845	0.389	0.996	2	Unreliable due to low sample
3	C-FR06001940_I6 C-FR06001940_RI	23	154	0.148	2.245	0.0741	2.947	0.952	2	Unreliable due to low sample
4	C-FR06001940_I6 C-FR06001940_RI	22	160	0.139	2.105	0.079	1.259	0.991	2	Unreliable due to low sample
5	C-FR06001954_I6 C-FR06001954_RI	54	393	0.136	2.067	0.0805	0.803	0.994	2	Unreliable due to low sample

The meanings of the columns are as follows:

Number: This is the replicate set number.

Replicate Set Names: self-explanatory.

GOI Average: This is the absolute number of molecules of DNA of GOI averaged across all the wells in the replicate set.

Control Average: This is the absolute number of molecules of DNA of the CNV Control averaged across all the wells in the replicate set.

Raw Ratio = GOI Average / Control Average (note that this ratio has systematic error because neither the GOI or the Control have been calibrated for absolute quantification)

Adjusted Ratio = Raw Ratio / Adjustment factor This number should be close to an integer (Adjustment factor = 0.0655 in this case)

CV of Ratio = coefficient of variation of Adjusted Ratio (CV = standard deviation / Adjusted Ratio)

Z-score: Number of standard deviations between the Adjusted ratio and the closest integer

$$\text{Z-score} = |\text{closest integer} - \text{Adjusted Ratio}| / \text{CV}$$

For the first row data we get: $\text{Z-score} = |2 - 1.922| / 0.0865 = 0.933$

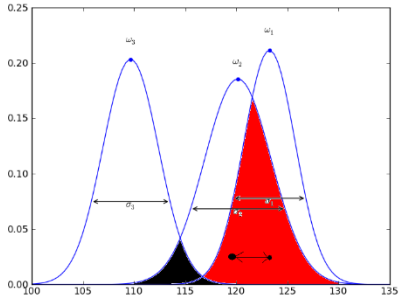
Confidence: Gives the statistical confidence that the reported CNV integer is correct. More details are given below.

CNV Call: The is the closest integer to the Adjusted ratio.

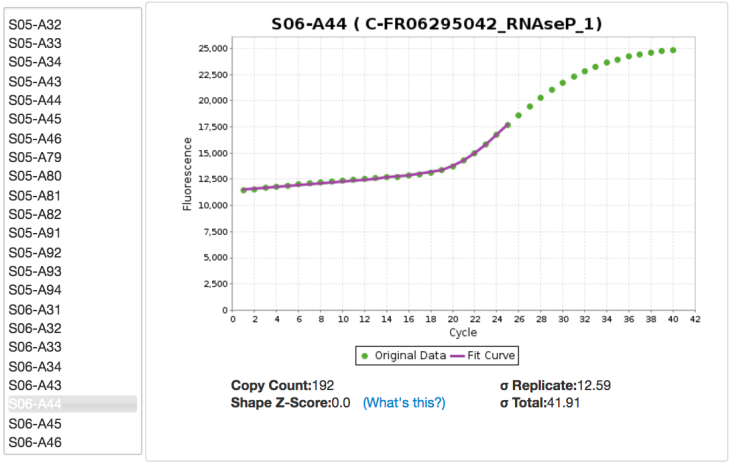
Comments: In the Table of CNV results, the rightmost "Comment" column gives information about samples that are unreliable, or ambiguous CNV calls (confidence < 0.75). Numerous experiments have shown that essentially all of the CNV calls with Confidence > 0.75 were experimentally verified to be correct. Note that PCR data acquired with very low amount of sample DNA (below 100 molecules) are not as reliable as datasets collected with higher DNA amounts, so comments to that effect are also displayed when appropriate.

More Information about the "Confidence" calculation. The Confidence score asks a simple question: When we observe a CNV value of 1.922 with CV = 0.0865, how confident are we that the actual CNV is 2 rather than 1? In this case, the Table reports "99.3%" Confidence (so we would expect to make an error only 0.7% of the time). The confidence score is based upon the analysis of the Gaussian distributions for means of 1, 1.922, and 2 with their respective uncertainties (see figure

below). The overlaps of these distributions (areas under curves computed using the cumulative distribution functions) is used to compute the confidence score.



CopyCount also allows you to see the raw data and fitted curves for each PCR reaction. Just select the well number that you want to observe and the curve will be shown.



CopyCount provides detailed information about the replicate sets and how they were averaged. See the Table below:

Replicate Sets

Table of results for each of the replicate sets: The "Mean DNA Copies" is the average over all the replicates after removing outliers. σ relative is the standard error in the mean and represents the reproducibility of the whole replicate set (with errors included from all sources including fitting and pipetting). σ relative is the appropriate error to use for comparisons of different samples of the same assay (i.e. the calibration error is not important for relative comparisons). For example, if the mean is 3000 copies and the σ mean is 100, then the 95% confidence interval (i.e. 1.96 standard deviations from the mean) is 2804 to 3196 copies. σ absolute is the standard error in the mean with inclusion of both σ and σ calibration. σ absolute is the appropriate error to use if you are comparing copy count measurements of two replicate sets that are for different assays.

Click here for more information about how to interpret your results.

Name	Wells in Set	Outlier Wells	Mean DNA Copies	σ Relative	σ Absolute
C-FR06001916_I6_1	S78-A31,S78-A32,S78-A3...		45.38	2.087	9.312
C-FR06001916_I6_2	S77-A31,S77-A32,S77-A3...		44.13	1.202	8.906
C-FR06001916_RNAseP_1	S78-A43,S78-A44,S78-A4...		358.1	8.254	72.10
C-FR06001916_RNAseP_2	S77-A43,S77-A44,S77-A4...		340.3	8.954	68.64
C-FR06001940_I6_1	S54-A31,S54-A32,S54-A3...		22.75	1.013	4.661
C-FR06001940_I6_2	S53-A31,S53-A32,S53-A3...		22.13	0.875	4.511
C-FR06001940_RNAseP_1	S54-A43,S54-A44,S54-A4...		153.8	4.439	31.07
C-FR06001940_RNAseP_2	S53-A43,S53-A44,S53-A4...		159.5	4.367	32.20
C-FR06001954_I6_1	S42-A31,S42-A32,S42-A3...		53.50	2.854	11.07
C-FR06001954_I6_2	S41-A31,S41-A32,S41-A3...		54.38	2.283	11.11
C-FR06001954_RNAseP_1	S42-A43,S42-A44,S42-A4...		392.8	9.059	79.07
C-FR06001954_RNAseP_2	S41-A43,S41-A44,S41-A4...		380.8	9.367	76.72

Note that Outlier wells are automatically thrown out. For example, S45-A91 (CopyCount=644 molecules) whereas the other 7 replicates are: 505, 471, 485, 464, 427, 464, 452 for an average = 466.9 with relative error = 9.293 molecules.

CopyCount gives the Assay information is given at the bottom of the page:

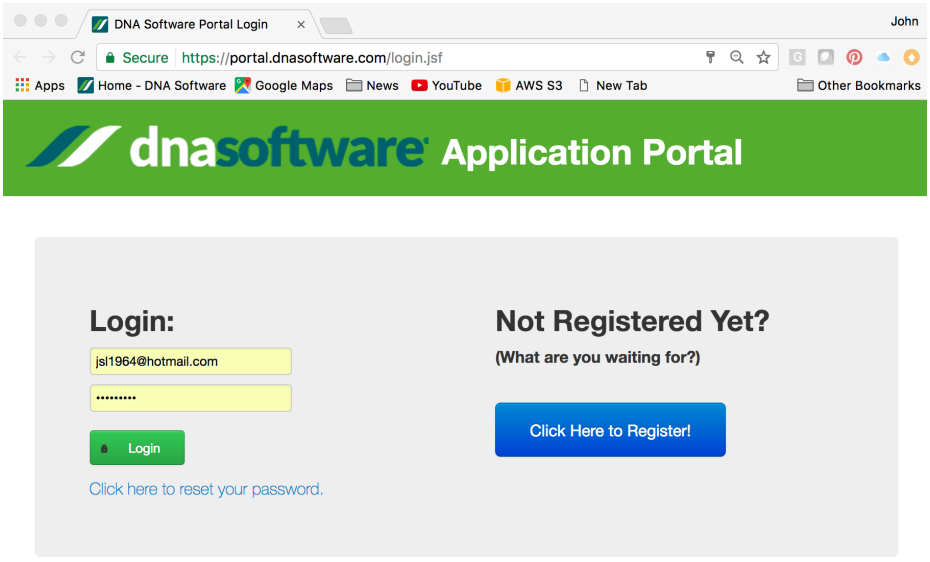
Assays

Name	Calibrated Assay ?	Double-Stranded Target	Amplicon Length	Detection Method	Primer Concentration	Probe Concentration	Using MGB
I6 new	false	true	71	TAQMAN	9.0E-7 M	2.0E-7 M	true
RNaseP_2	false	true	87	TAQMAN	9.0E-7 M	2.0E-7 M	false

Appendix: Registration

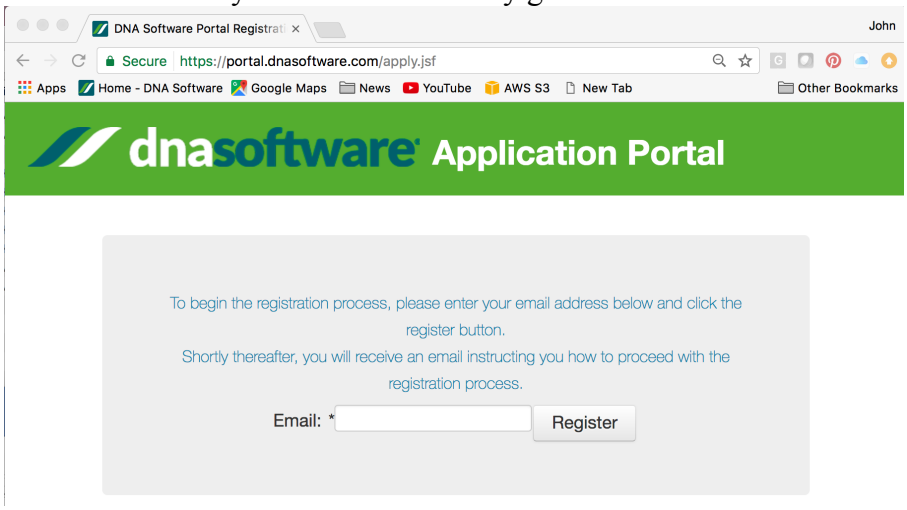
Registering for the DNA Software Application Portal is painless and mostly self-explanatory, but details are provided here if you need them. If you have any problems with registration, please contact Joseph Johnson at: Joseph.johnson@dnasoftware.com

Step 1: Go to the DNA Software Application Portal: https://portal.dnasoftware.com/login.jsf

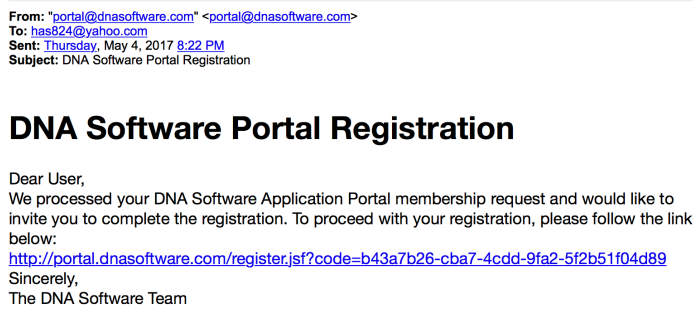


To register, just Click on the BLUE button "Click Here to Register".

Provide your email address and then you will automatically get an email sent to that address.



Step 2: The email you will receive will look like the screenshot shown below:

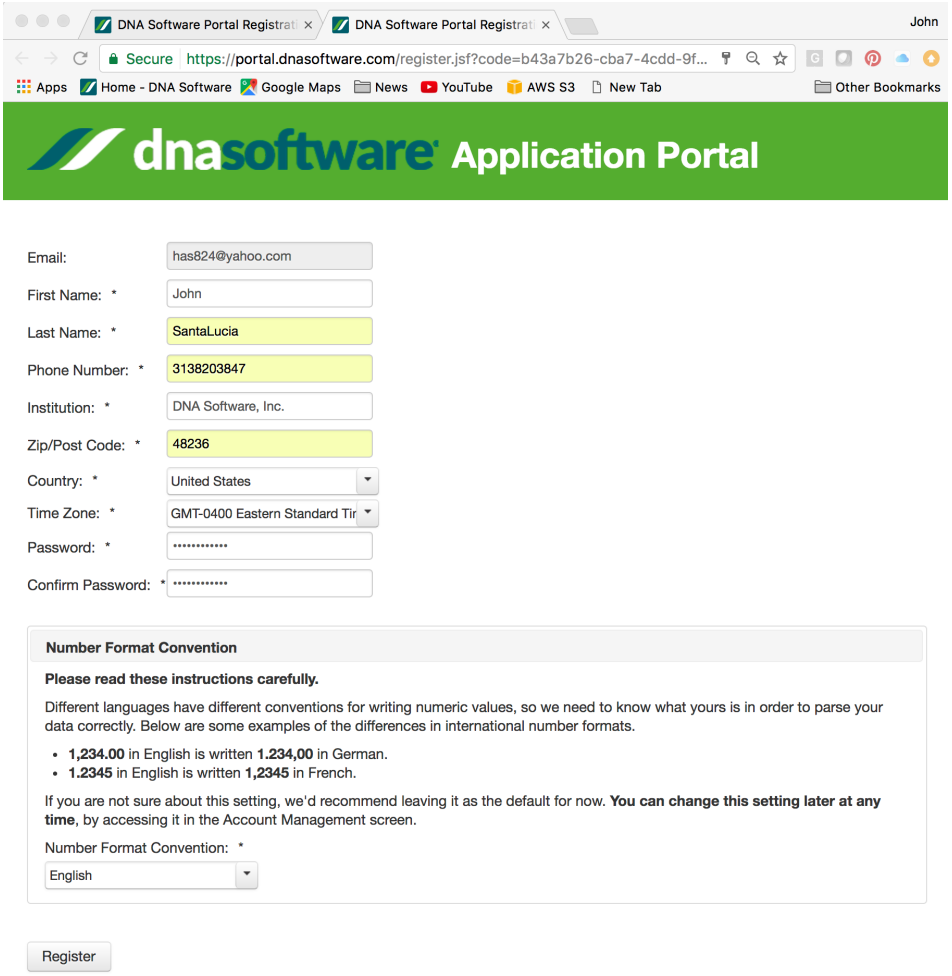


From: "portal@dnasoftware.com" <portal@dnasoftware.com>
To: has824@yahoo.com
Sent: Thursday, May 4, 2017 8:22 PM
Subject: DNA Software Portal Registration

DNA Software Portal Registration

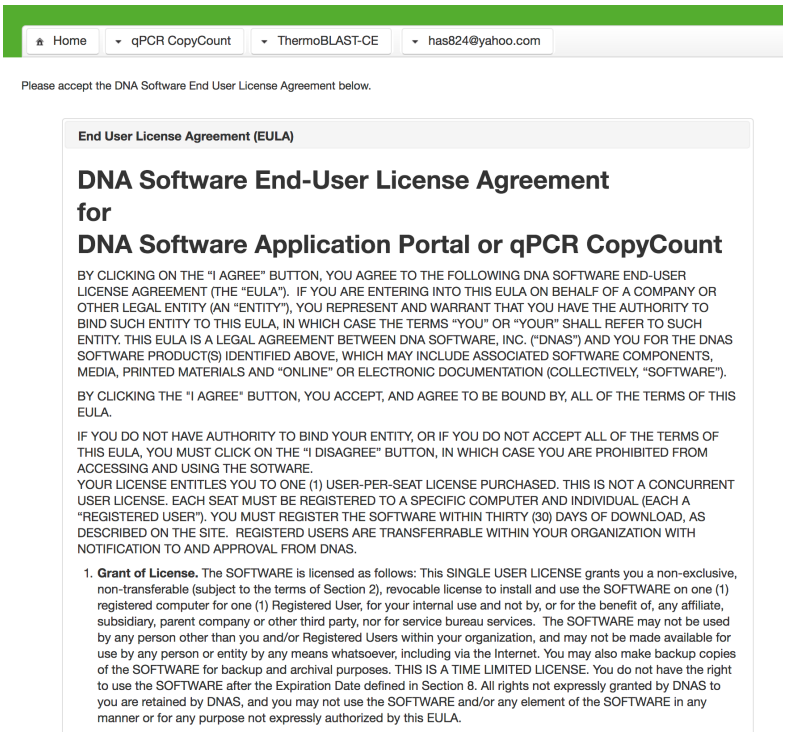
Dear User,
We processed your DNA Software Application Portal membership request and would like to invite you to complete the registration. To proceed with your registration, please follow the link below:
http://portal.dnasoftware.com/register.jsf?code=b43a7b26-cba7-4cdd-9fa2-5f2b51f04d89
Sincerely,
The DNA Software Team

Step 3: Click on the link in the email and that will take you to the Registration web page shown below. Complete the registration. Once you and done filling out the form, Click the "Register" button.

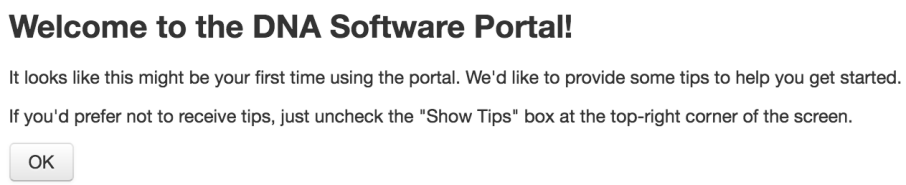


Step 4: That's it! You are now registered to use the DNA Software Application Portal. If you have already paid for the license to use the software, then you are all set and you can begin using the software immediately. If you have not paid, then DNA Software staff will contact you to discuss license options.

The first time you log on to the account, you will be asked to Accept both the "End User License Agreement (EULA)" and the "Privacy Policy". On each page, scroll to the bottom and click the "Accept" button on the Left side.



Once you have accepted the EULA and Privacy terms, then you will see a popup message:



Click the "OK" button and then you will be taken to the Application Portal where you can immediately start submitting your jobs.