

Comparison of Normalized Values in GeneSpring GX 7.3.1 and GX 9.0.5



Agilent Technologies

Comparison of normalized data values in GX7 and GX9

Dataset used: We used the GSE 6711 series from GEO to carry out this analysis.

First, we carried out an experiment using this dataset in both GX7 and GX9, using the default settings provided in both these versions.

Default steps followed to import Agilent one-color data into GX7 (as described in the saved Agilent 1-color scenario):

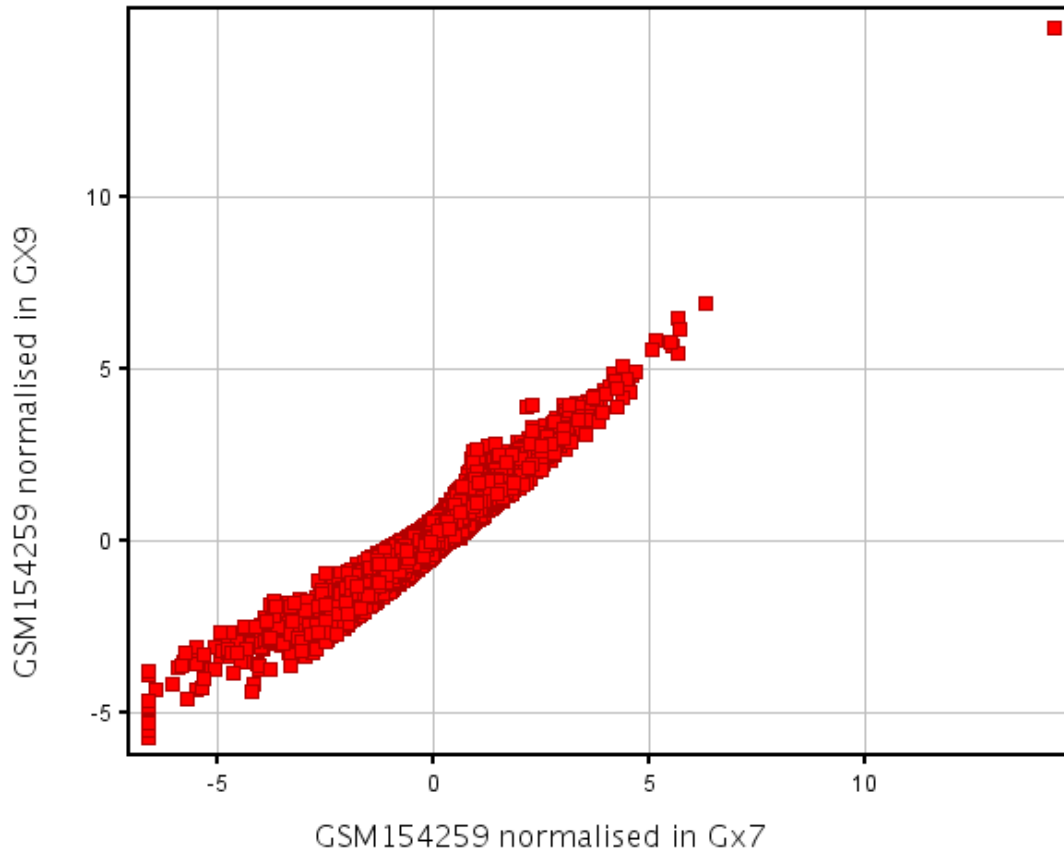
1. Use spot information to flag the data, using default Agilent parameters.
2. Summarization.
3. Threshold to 5.
4. Per Chip Scaling to 50th Percentile.
5. Per Gene median normalization.

Default steps followed to import Agilent one-color data into GX9:

1. Use spot information to flag the data, using default Agilent parameters.
2. Summarization.
3. Threshold to 5.
4. Median Shift Normalization to 50th percentile. (Same as step 4 in the GX7 workflow)
5. Baseline Shift to median of all samples. (Same as step 5 in the GX7 workflow)

Please note that, between steps 3 and 4 in the GX9 workflow, a logarithm to base 2 is performed on the data. When comparing GX7 and GX9, we logged the GX7 data at the end of the process, to compare the 2 sets of data.

After carrying out the above mentioned steps in GX7 and GX9 on 6 samples, the normalized data values for all samples were exported out and plotted against each other. This plot showing the values for the first sample in Gx7 vs. values for first sample in GX9 is reproduced below:

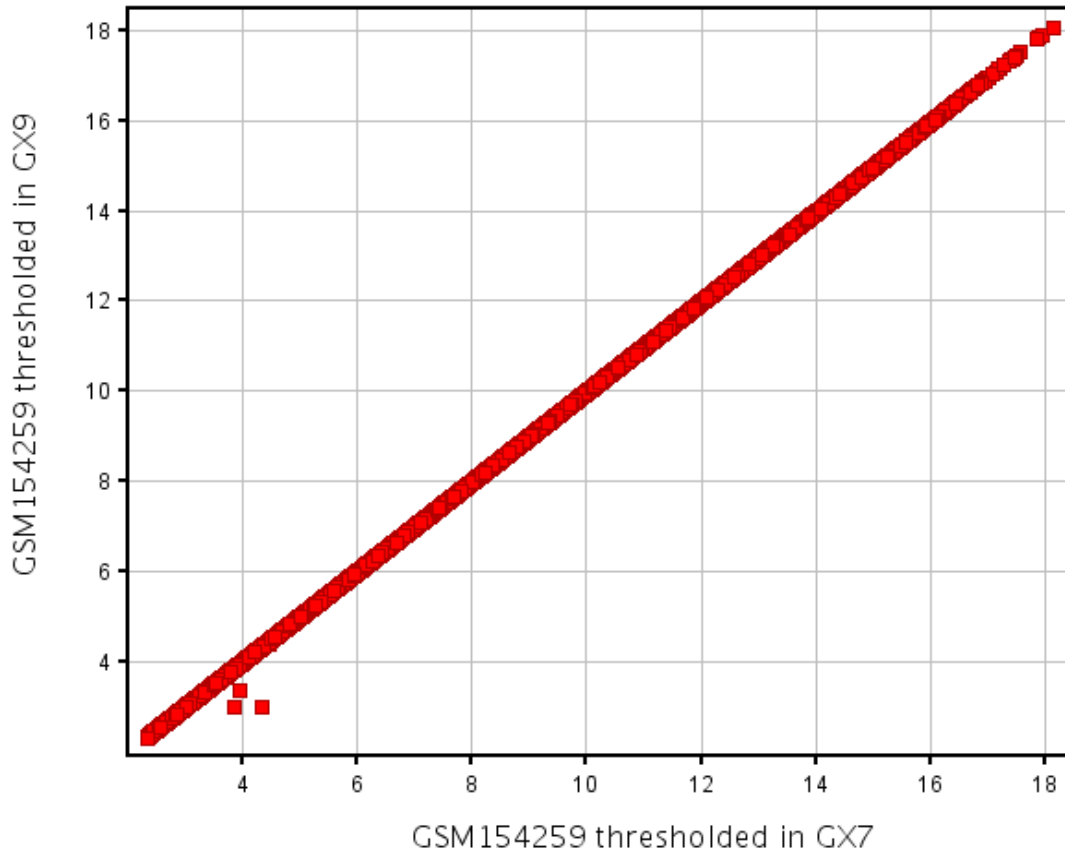


As seen above, there is a fair bit of scatter in this graph around the $y = x$ line, meaning that the normalized values from GX7 and GX9 do not match exactly.

We then compared normalized data values out of GX& and GX9 **at every step** in the normalization procedure, to pinpoint the exact reasons for this disparity.

At step 1:

The following graph plots GX7 data versus GX9 data, immediately after thresholding to 5 (no further normalization):



In this step, data values below 5 are pushed up to 5. As seen above, all points except 3 all on the $y = x$ line, i.e. all probes except those 3 have the same numerical value after thresholding in GX7 and GX9.

Consider one of these 3 probes, with probe ID A_23_P6335. Its value in GX7 is 20.25455 in the linear scale, (or 4.340174 in the log scale), whereas in GX9 it is 2.9940972 in the log scale. In order to understand how these values came to differ, we look at a table showing all replicate probes, their signal and relevant flag values, for one of these 3 probes.

ProbeName	gProcessedSignal	glsPosAndSignif	glsWellAboveBG
A_23_P6335	3.876639	0	0
A_23_P6335	10.16508	1	0
A_23_P6335	20.25455	1	1
A_23_P6335	7.285747	1	0
A_23_P6335	10.14092	1	0
A_23_P6335	7.406694	1	0
A_23_P6335	2.877337	0	0
A_23_P6335	6.167483	1	0
A_23_P6335	8.481737	1	0



A_23_P6335

3.017158

0

0

* gIsPosAndSignif: flag indicating whether signal is positive and significant

** gIsWellAboveBG: flag indicating whether signal is above background

As seen above, only one out of the 10 replicate probes could be considered positive and significant as well as well above background. Now, in GX7, we consider this spot information for flagging of data and use it to filter out probes before summarization, and then summarize based on the present probes. Therefore, in this case, *all except for one of the replicate probe values will be ignored during summarization of data across replicate probes*. Hence, the value assigned would be 20.25455. Conversely, in GX9, we have taken the decision to include ALL probe values during summarization across replicate probes, because it was felt that ignoring probe values that reflect a below background situation would lead to artificially high signal values. Hence, the signal value for this probe in GX9 after summarization would be:

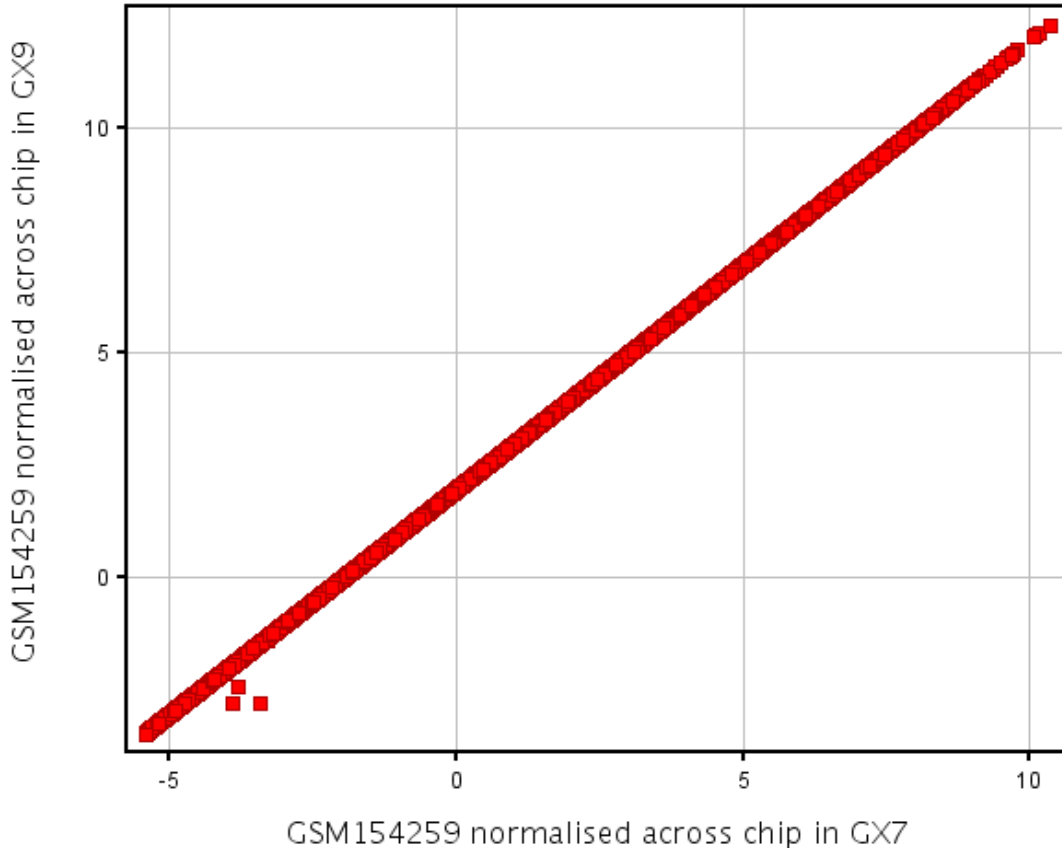
$$(3.876639 + 10.16508 + 20.25455 + 7.285747 + 10.14092 + 7.406694 + 2.877337 + 6.167483 + 8.481737 + 3.017158)/10 = 7.967334$$

which is 2.9940972 in the log scale.

In addition to filtering out absent probes based on flags, GX7 also thresholds all data to 0.01, whereas, as explained above, in GX9 we chose not to threshold before summarization so as to reflect the actual data values.

At step 2:

The following graph plots GX7 data versus GX9 data, after thresholding to 5 and per chip normalization to the 50th percentile:



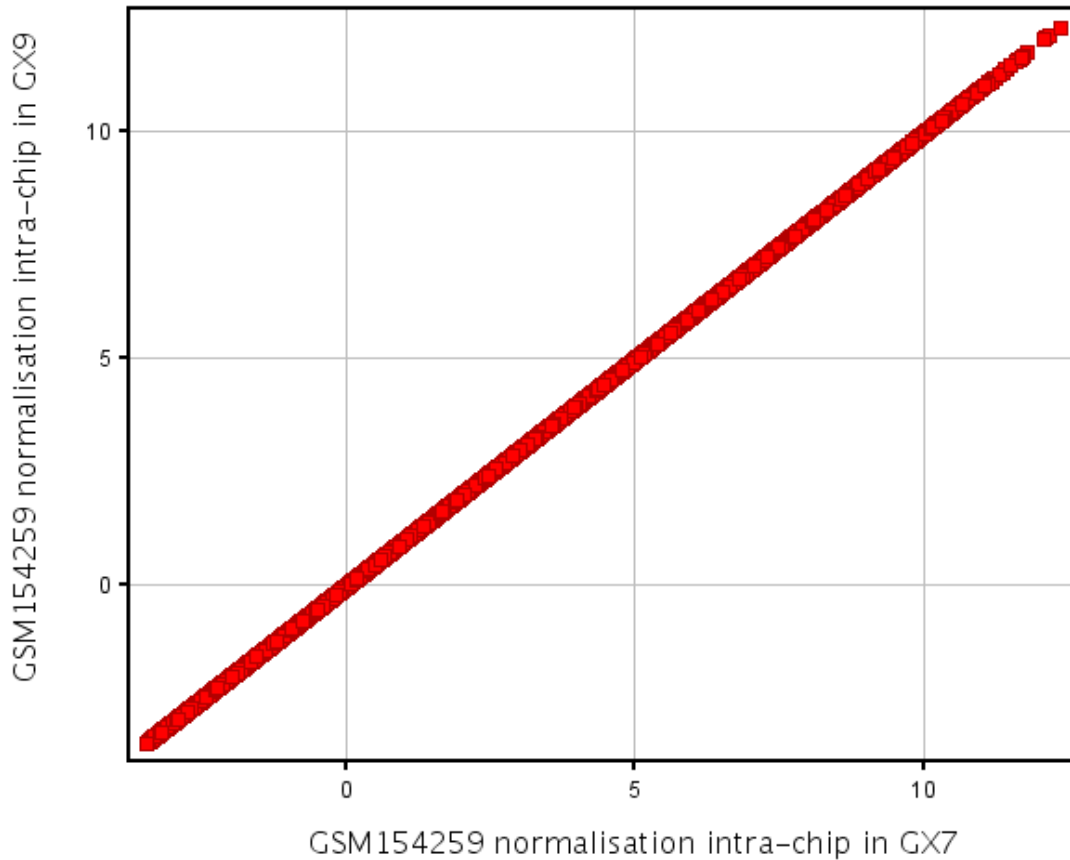
In this step, the median of all probe values within a sample is calculated and the data is modified on the basis of this median, such that, after this processing, the distribution of all values becomes a normal distribution.

The above plot shows, again, all but the same 3 probes falling on the same straight line: $y = x + 1.9866$. Thus, there is a *shift* of 1.9866 in the data between GX7 and GX9.

Now, during normalization to 50th percentile within a chip, GX7 uses only the probes marked as present to calculate the median of signal values within a chip. GX9, on the other hand, uses ALL probes to calculate the median. In this particular case, a significant number of probe values would be considered 'absent' and therefore would not be considered in the median calculation, because the default flag import settings^{***} would mark them as such. Hence, the median calculated by GX7 turns out to be ~7.771 (in the log scale), whereas the median calculated in GX9 is ~5.771. This will lead to a difference of ~2 in the log values between GX7 and GX9.

Confirmation:

In order to confirm the varying contribution of control probes and absent probes in the normalization process in both versions, we decided to neutralize the effect of absent probes by NOT using spot information to flag the probes, i.e., by considering ALL probes as present. We then carried out all the normalization steps except for per gene median normalization (baselining). The corresponding comparison scatter plot between normalized values (without baselining) in GX7 and GX9 was as follows:



Thus, if ALL probe values were considered in both versions, to calculate the median of values within a sample, and then median shift normalization was carried out, then the GX7 and GX9 data matched perfectly. All points are now on the $y = x$ line.

*** Default Flag Import settings in GX9 are shown in the figure below. The default settings are the same in GX7 and GX9, except for an additional spot problem "Flag is not found" being included in GX9.

New Experiment (Step 2 of 3) ✖

Advanced Flag Import
Advanced Flag Import Settings.

Use spot information in data files to flag the data

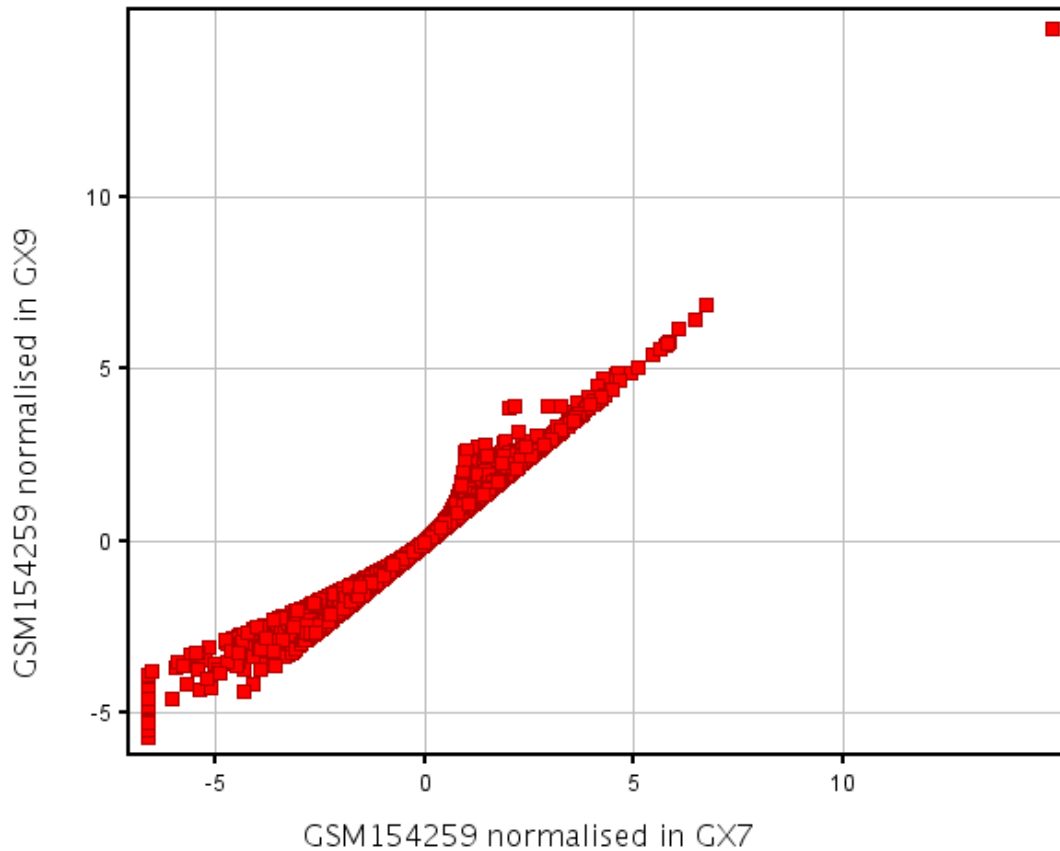
Include background reading in flag settings

Spot Problems

	Present	Marginal	Absent
Feature is not positive and significant	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Feature is not Uniform	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Feature is not above background	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Feature is Saturated	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Feature is manually marked	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Feature is not found	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Feature is a population outlier	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Background is not uniform	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Background reading is a population outlier	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

At step 3:

This data underwent all the steps outlined as the default sequence of steps, i.e. thresholding to 5, shifting the median to 50th percentile (with no flagging, i.e. all probes present), and baseline transformation for each gene across the 6 samples.



The scatter seen in this plot is because, in GX7, the values are retained in the linear scale, whereas GX9 works with values in the log scale. This is illustrated below:

In GX7:

Before Baseline transform, the values for one probe across 6 samples looks like this:

Probe Name	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
A_23_P391586	435.0209	376.901	353.7839	15.07799	8.447025	7.053943

The median here would be $(15.07799 + 353.7839)/2 = \mathbf{184.430945}$. All the individual values are now **divided** by this value. The data now is:

Probe Name	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
-------------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------



A_23_P391586 2.35872 2.043589 1.918246 0.081754 0.0458 0.038247

Or, in the log scale:

Probe Name	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
A_23_P391586	1.238004	1.031105	0.939788	-3.61257	-4.44849	-4.70851

In GX9:

Before Baseline transform, the values for the probe across 6 samples looks like this:

Probe Name	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
A_23_P391586	8.755012	8.552899	8.461639	3.910273	3.074112	2.816872

Please note that these are in the log scale. The corresponding linear values are almost exactly equal to the linear values used in GX7.

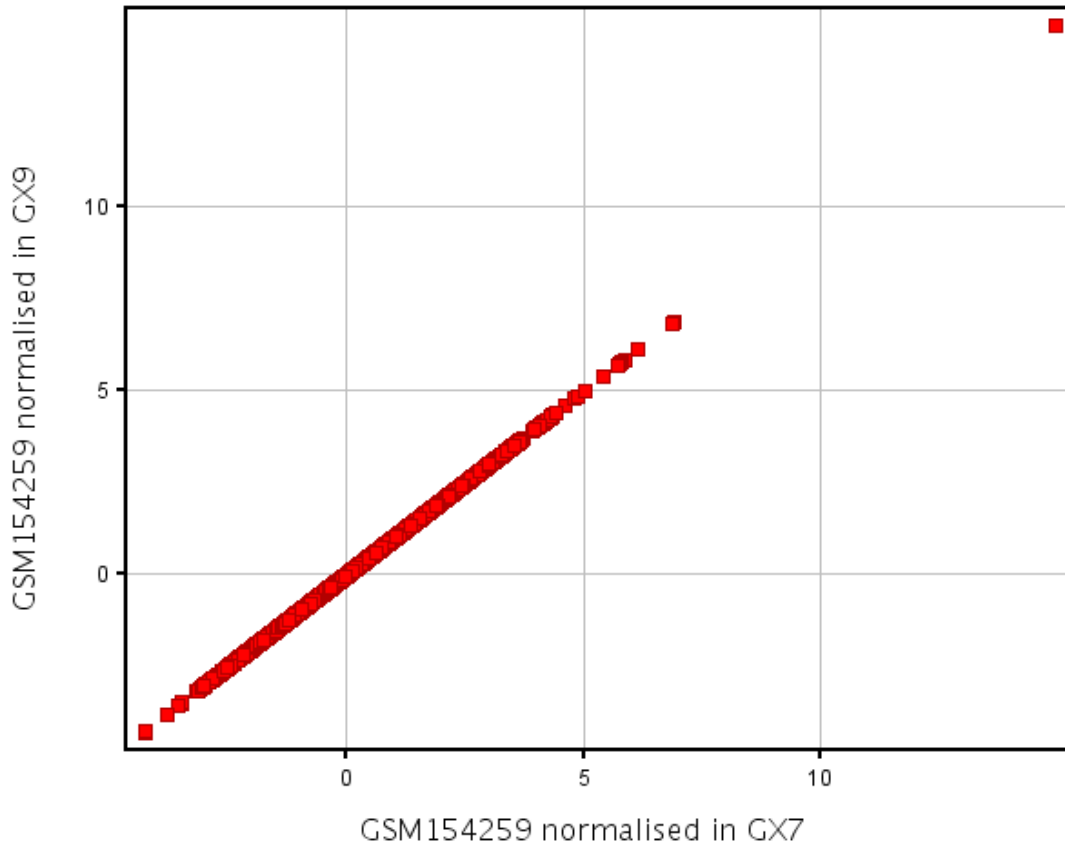
The median here would be $(3.910273 + 8.461639) / 2 = 6.185956$. This value is **subtracted** from each of the individual values. The data now is:

Probe Name	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
A_23_P391586	2.569056	2.366943	2.275683	-2.27568	-3.11184	-3.36908

Now, subtracting a number *a* from another number *b* in the log scale is equivalent to dividing *b* by *a* in the linear scale. However, in this case, because the median calculations are carried out in different scales, the median value arrived at is *not the same number*. The median calculated in GX7 was 184.430945. This number is **~7.5269** in the log (to the base 2) scale, whereas the median arrived at in GX9 was **6.185956**. Thus, the median values considered in each case are different, resulting in a difference in the final normalized values.

Confirmation:

We took an odd number of samples (five), neutralized the effect of absent probes as above, and then carried out all normalization steps including per gene normalization to median. Having used an odd number of samples, the same value will be taken as median both times, and the scale will not matter. The corresponding scatter plot between normalized values in GX7 and GX9 was as follows:



As this plot illustrates, the scatter in the previous plot is a result of the difference in the median calculated across samples, when there are an even number of samples, as a result of the different scales being used.

Summary of findings:

The discrepancy in the values shown in figure 1 can be attributed to the following:

1. In the summarization step, GX7 raises all values below 0.01 to 0.01. Thus, while GX9 takes all probes, GX7 ignores negative and lower probe values. Negative probe values reflect a below background situation and ignoring these leads to artificially high signal values, hence we made a decision to include these in GX9.
2. Absent probes are ignored in GX7 during per chip normalization.

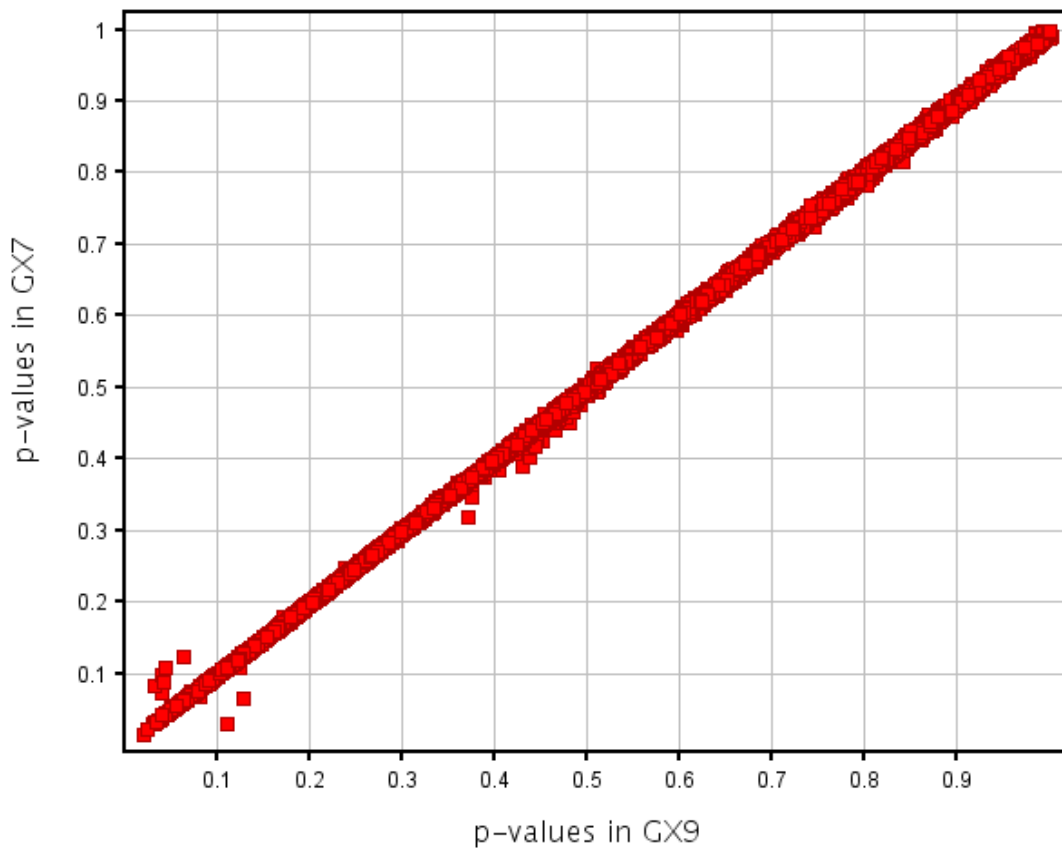
During per gene normalization (baselining across samples), the data values are in log scale in GX9, whereas in GX7 they are in linear scale. This results in a difference in the median value which is considered for baselining. In addition, negative values are *ignored* during this step in GX7, and are pushed to 0.01. Only values above 0.01 are used in the calculation of the median, and only these values are then baseline transformed. Conversely, GX9 considers ALL values during baselining.

Effect of the differences in normalized values on statistical analysis:

Assuming that, in most cases, a user would prefer the default settings in the tool, we compared the results obtained from both versions using the default workflow and flag settings. We carried out the same steps in GX7 and GX9 as outlined at the beginning (thus getting Figure 1 when we compare the resultant normalized values in GX7 and GX9), and then performed statistical analysis.

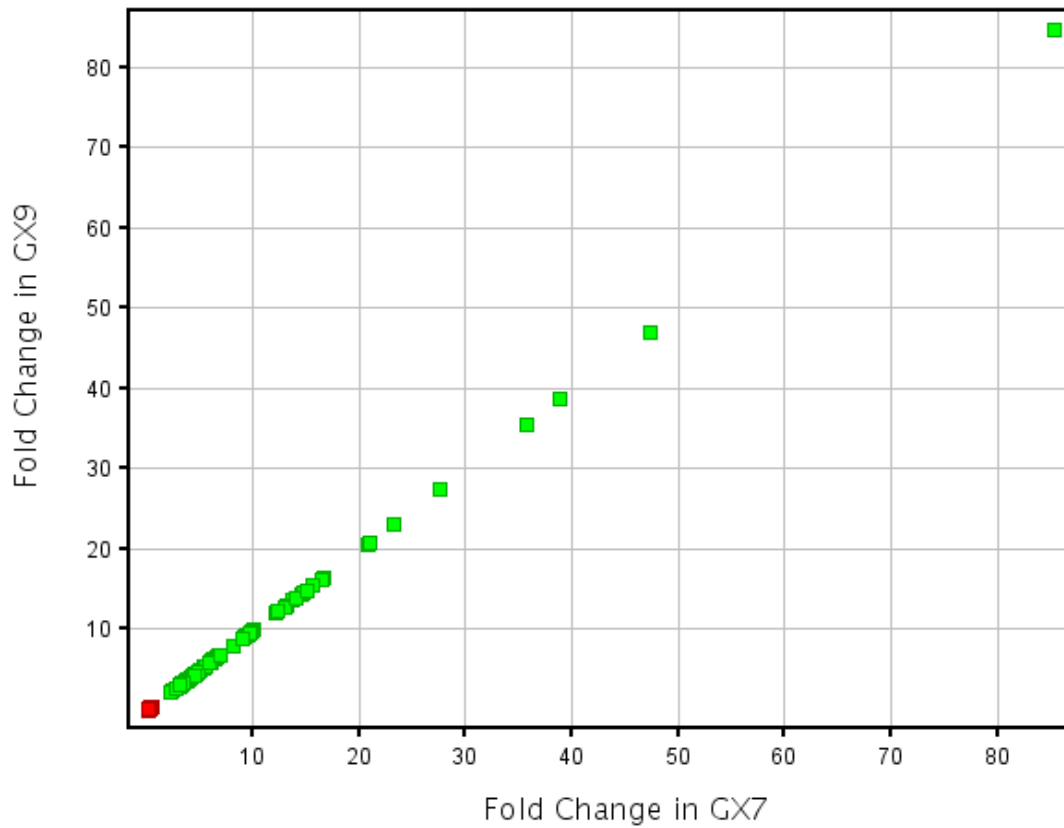
Equivalent statistical tests were performed. In GX7, a parametric t-test with unequal variance was carried out, and then Benjamini Hocheberg FDR was used for multiple testing correction. In GX7, an unpaired t-test with unequal variance was carried out, and then Benjamini Hocheberg FDR was used for multiple testing correction.

The p-values obtained for all probes from GX7 and GX9 are plotted against one another. As seen below, the correlation is very high:



Thus, the results of a t-test carried out on data normalized in GX7 and GX9 would yield almost similar results.

Similarly, if we plot the fold changes for statistically significant probes, obtained from GX7 and GX9, we get the following scatter plot:



Thus, fold change for probes is also very highly correlated, when you compare values from GX7 and GX9.