

INTERNSHIP REPORT

Detection of DNase-I Hypersensitive sites at the human KIBRA gene locus by Quantitative PCR

An internship report presented in partial fulfillment of the requirement of the Professional Science Master's in Computational Biosciences

Krupa Arun Navalkar
Computational Biosciences Program
Arizona State University

Dr. Jeffrey Touchman, Ph.D.
Internship Advisor
School of Life Sciences
Arizona State University

Dr. Scott E. Bingham, Ph.D.
Dr. Jackiewicz Zdzislaw, Ph.D.
Committee Members

Technical Report Number: 07-15

Acknowledgement

I would like to acknowledge and thank Dr. Jeffrey Touchman, for his support and guidance and for giving me the opportunity to work on this project. I am grateful to Dr. Scott Bingham and Dr. Zdzislaw Jackiewicz for agreeing to be on my Committee and providing me with their invaluable feedback.

I would like to thank the staff at Dr. Touchman's Lab in Translational Genomics (TGen) for creating an environment where students can work with other scientific officers and for trusting me to join their team. I would also like to thank my colleagues at TGen – Translational Drug Development (TD2), Karen Fisher, Heather Taylor and Ahmet Kurdoglu for sharing their experimental data with me.

Lastly, I would like to thank Mrs. Renate Mittelmann, my parents and friends....for always being there.

Index

1. Abstract.....	5
2. Introduction.....	6
3. Methods and Discussion.....	11
4. Results.....	21
5. Future Work - Recommendations.....	25
6. References.....	27
7. Appendix-I: Cell lines,Cell culture methods & PCR Protocol.....	30
8. Appendix-II: 384 well PCR plate layout.....	35
9. Appendix-III: Results from preliminary experiment.....	35
10. Appendix-IV: Known LCR list with EST data to be used as positive controls for microarray approach.....	35

List of Figures

Figure 1: General outline of the Protocol (Dorschner et. al. 2004).....	9
Figure 2: Amplification Plot as seen in ABI Prism 7900 SDS software...11	
Figure 3: Quality Control for Primer rows A-D.....12	
Figure 4: Quality Control for Primer rows E-G.....12	
Figure 5: Dissociation Curve as seen in ABI Prism 7900 SDS software...13	
Figure 6: Snapshot of clipped data from delta Rn vs. cycle number graph.....	15
Figure 7: Export data from excel into LinRegPCR software.....16	
Figure 8: ‘All Sample View Screen’ in LinRegPCR software.....17	
Figure 9: Selecting Data Points for each individual curve in LinRegPCR.....	18
Figure 10: Final Results output sheet with PCR efficiencies from LinRegPCR.....	19

Abstract

The Kibra gene is located on human chromosome 5 at location: 5q34-q35.2. It produces a 1,113 amino acid long protein that has a molecular mass of about 125 kilodaltons. This phosphoprotein contains two N-terminal, conserved tryptophan domains (i.e., WW) around 40 amino acids long and a glutamic acid rich stretch in its C-terminal half. It is a protein found to be expressed in breast cancer cells and memory related brain structures such as the hippocampus and temporal lobes.

It has long been known that the non-protein coding regions of DNA are involved in the transcriptional control of a specific gene. These non-protein coding regions of DNA that either increase or decrease the expression levels of a certain gene are also known commonly as 'Locus Control Regions'. The objective of this research project is to identify such functional, non-coding elements that regulate the transcription of the Kibra gene.

These non-coding elements are often sites in the genome that are sensitive to the enzyme DNase-I. The project aims to precisely localize such cis-regulatory sequences and other functional modalities encoded by these DNase-I Hypersensitive sites (traditionally localized using Southern blot) through the use of the quantitative polymerase chain reaction approach in cancer cell lines.

Introduction

The Kibra gene is located on human chromosome 5 at location: 5q34-q35.2. It produces a 1,113 amino acid long protein that has a molecular mass of about 125 kilodaltons. This phosphoprotein contains two N-terminal, conserved tryptophan domains (i.e., WW) around 40 amino acids long and a glutamic acid rich stretch in its C-terminal half. ^[31] WW motif containing proteins are known to mediate protein-protein interactions which are essential for the transmission of cellular signals. The Kibra is of interest because it is seen to be highly expressed in the brain and might therefore play a role in specific cognitive pathways.

It is a protein found to be expressed in breast cancer cells in relation to its interaction with the Dynein Light Chain 1(DLC1) ^[33] protein leading to growth stimulation in the same. It is also found to be expressed in its truncated form, where its missing the first 223 amino acids in memory retrieval related brain structures such as the hippocampus and temporal lobes, and functional magnetic resonance imaging (fMRI) studies show a significant role of Kibra in memory performance. ^[32]

The genetic material in humans is packaged in the form of chromatin within the nucleus. It is basically a tightly wound complex of DNA and histone proteins. The expression of various genes in different tissues at specific time points in the cell cycle is dependent upon which sections of chromatin are condensed (heterochromatin) or open (euchromatin) during that specific cellular phase. Genes in the euchromatin are exposed to transcriptional mechanism and this is seen mainly of most of the active genes. Genes in the heterochromatin do not get expressed as they are not exposed to the cells transcriptional machinery. Thus gene expression is tightly associated with the structure of the chromatin in that region.

How and why then, this folding and unfolding of chromatin occurs and is timed within a cell cycle is not yet fully understood. There are several theories ^[20] based on experimental

data to explain how and why chromatin structure unfolds, which suggest various techniques like remodeling complexes causing histone modifications (acetylation, methylation or phosphorylation of histone amino acids) to epigenetic inheritance leading to maintenance of gene expression patterns throughout generations.

It has long been known that the non-protein coding regions of DNA are involved in the transcriptional control of genes, as they are often capable of binding proteins that modulate transcription. These non-protein coding regions of DNA that either increase (enhancer or promoters) or decrease (repressors) the expression levels of a certain gene are also known commonly as ‘Locus Control Regions’ [9,10,17,19,26]. One of the most well studied Locus Control regions is the human HBB (β -globin) LCR [11]. The objective of this research project is to identify functional, non-coding elements that are involved in the regulation of the Kibra gene so as to understand its possible contribution in a disease process like memory loss or performance.

Non-coding elements are often sites in the genome that are sensitive to the enzyme DNase-I, which attacks unprotected or unbound (to any proteins) DNA. These nuclease hypersensitive sites are said to lack the presence of a nucleosome, thereby, making them accessible and almost two-fold more sensitive to DNase-I activity [26] than the other regions of chromatin. An advantage of using DNase-I for detecting these sites is its relatively low sequence specificity. The main types of DNase-I hypersensitive sites are: constitutive, inducible, tissue specific and developmental [26]. The constitutive and inducible sites are often found upstream of the promoter region of a gene.

The project aims to locate such cis-regulatory sequences and other functional modalities corresponding to these DNase-I Hypersensitive sites (traditionally localized using Southern blot) through the use of the quantitative polymerase chain reaction approach using four cell lines. The cell lines used will be T-47D (breast cancer), U-87 MG (Glioblastoma, brain) and the MDA-MB-436 (breast tissue). One of these cell lines, IMR-90, is a normal lung fibroblast and along with the U-87 MG cell line is a low expresser of Kibra. The T-47D and MDA-MB-436 cell lines are high expressers of Kibra.

We expect different hypersensitivity profiles between the high-expresser and low-expresser cell lines.

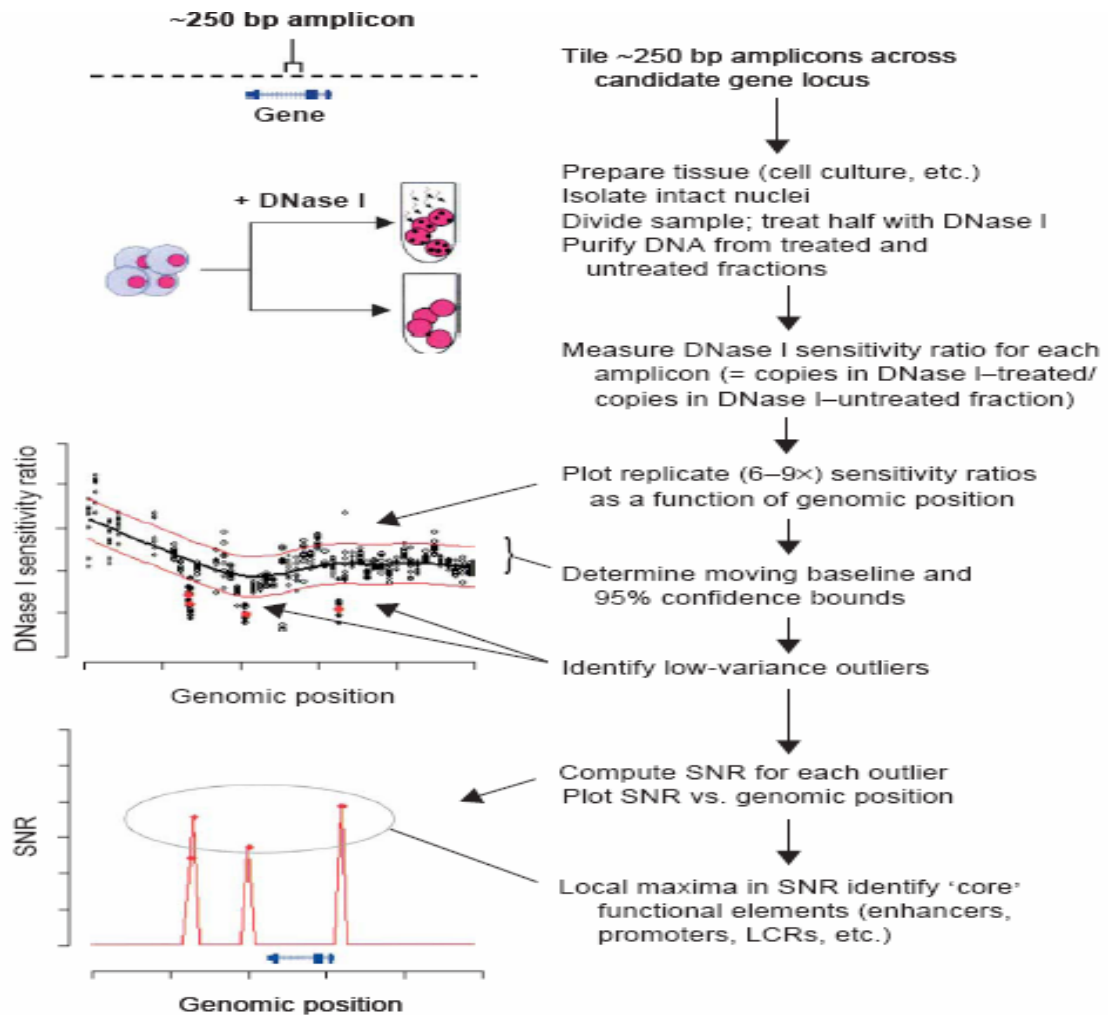
The $2^{-\Delta\Delta C_t}$ method [1, 4, 5]

The polymerase chain reaction can be performed either qualitatively or quantitatively. Two approaches for quantitative PCR include absolute quantification and relative quantification. In absolute quantification, the exact copy number of the PCR product is estimated using a standard curve. We used the relative quantification approach as we needed to compare the PCR amplification signal obtained from a DNase-I treated group of genomic DNA fragments versus, an untreated group spanning the genomic region of the Kibra gene.

The methodology used to analyze functional, non-coding, regulatory genetic elements is the $2^{-\Delta\Delta C_t}$ method [4, 5] which allows monitoring of the digestion of DNA fragments by DNase-I as it allows an easy comparison between the Cycle threshold values of the DNase-I treated versus those of the untreated fragments of genomic DNA. We would be using the quantitative polymerase chain reaction (QT-PCR) for this experiment and calculating the results from the QT-PCR experiment through the $2^{-\Delta\Delta C_t}$ method to obtain information on the Kibra gene locus.

In this method the cell lines will be harvested at 70% growth confluence levels and their nuclei will be isolated. Half will be treated with varying concentrations from 10-80 U/ml of DNase-I enzyme and the other half will not be treated with DNase-I. The purified DNA from such extracts will be subjected to quantitative polymerase chain reaction using SYBR green and an ABI 7900 instrument.

Figure 1: General outline of the Protocol (Dorschner et. al. 2004) ^[1]



Thirty primer pairs approximately 250 bp in length \pm 50 bases were designed across the Kibra gene region for the qPCR experiments. Individual primers have a temperature of melting (T_m) between $60^\circ\text{C} \pm 2^\circ\text{C}$ using the program Primer3.

We are in the process of conducting a preliminary experiment to test DNase-I hypersensitivity in known hypersensitive regions using two PCR amplicons designed for the same so as to arrive at a single concentration of DNase-I to be used for the entire experiment. These assays would then be used throughout the project as positive controls. Output from the quantitative PCR reaction includes Ct (cycle threshold) values and PCR efficiencies of each individual amplicon.

The Plate setup for the final PCR Reaction would be as follows:

(10-80 U/ml) DNase –I treated and untreated, 2 primers spanning known hypersensitive regions (eight replicates each).

(10-80 U/ml) DNase – I treated and untreated RHO primer (eight replicates each)

A detailed Excel worksheet containing the plate layout has been attached in Appendix-II of this report.

For absolute quantification in a typical quantitative polymerase chain reaction one needs to perform a standard curve experiment to validate the efficiency of a test reaction and the internal reference gene amplicon. Since this would be a relative comparison and not an absolute one, we would only be running a standard curve using the reference gene reaction as an internal control measure on every PCR reaction plate to verify the accuracy of the results. The reference gene reaction is RHO (Rhodopsin gene). Rhodopsin is present on Chromosome 3 of the human genome and is DNase-I insensitive and therefore, it is selected as the reference gene.

We would then measure the DNase-I sensitivity ratios of individual reactions versus their genomic location within the Kibra sequence and determine the moving baseline and 95% confidence bounds using R coding. We then identify the low-variance outliers and compute the signal to noise ratio for each outlier versus their genomic position within Kibra. These local maxima identified would be the putative locus control regions we are seeking.

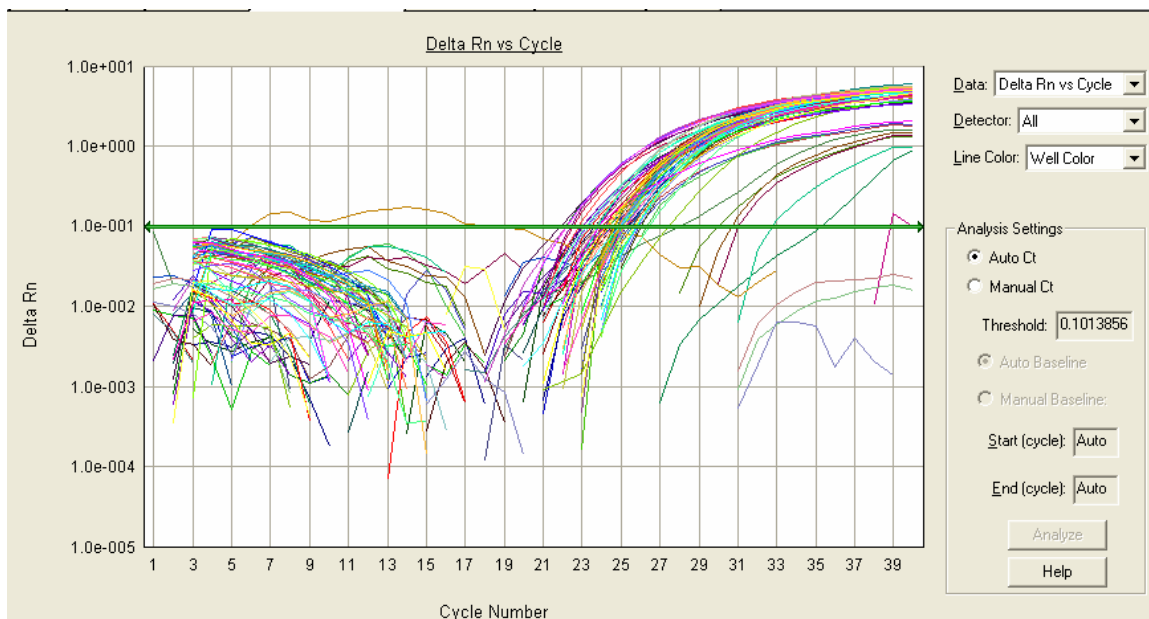
Once we arrive at a DNase-I concentration that shows the presence of DNase-I hypersensitive sites within these 2 primers, we could later use this same concentration for the microarray approach to detect hypersensitive sites in the different cell lines.

Methods and Discussion

The following are phases which we need to complete, to obtain relevant data through preliminary data analysis: ^[1]

- 1.) Obtaining cycle threshold (C_t) data for all replicates of the 2 test primers designed across previously known hypersensitive regions along with the Rhodopsin primer internal control C_t values.

Figure 2: Amplification Plot as seen in ABI Prism 7900 SDS software



- 2.) Amplification efficiency correction is done using the following criteria ^[5]:
(personal communication, Dr. Jan Ruijter)
 - a. The efficiency values must be within the median range of efficiency values ± 0.05
 - b. The range of R values is between 0.995 to 0.999 and,
 - c. The range of efficiency values is between 1.8 and 2.0

3.) The melting curve analysis was done using the ABI Prism 7900 SDS software, so as to ensure that every amplicon showed a single product curve. Quality control for each amplicon has also been completed by running them on a polymerase gel to see that every primer pair gives only a single band i.e. a single PCR product.

Figure 3: Quality Control for Primer rows A-D

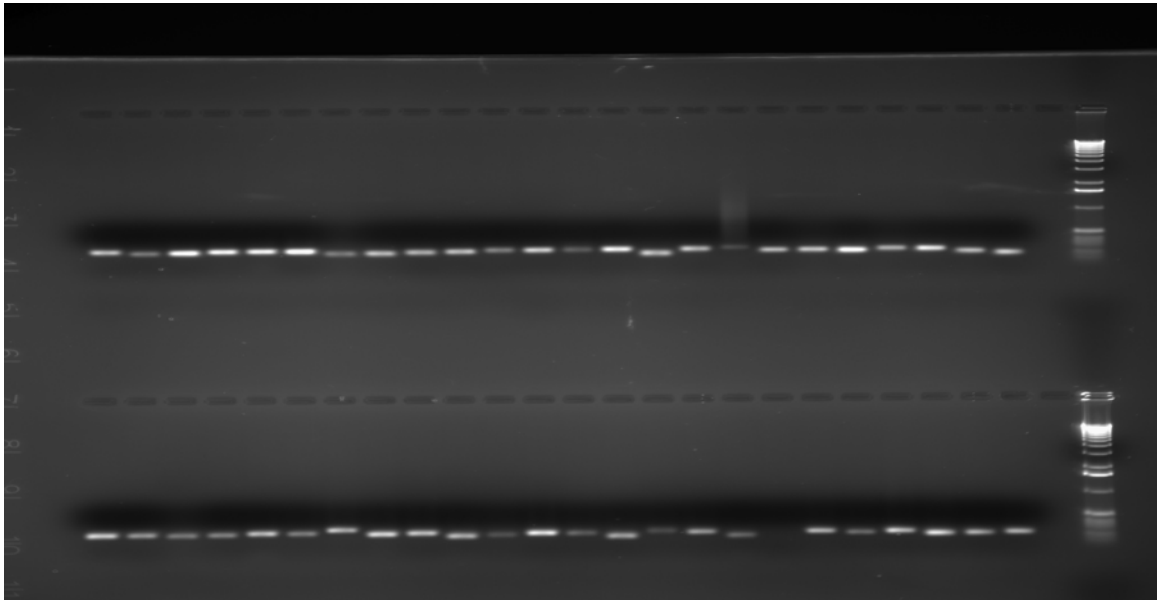


Figure 4: Quality Control for Primer rows E-G

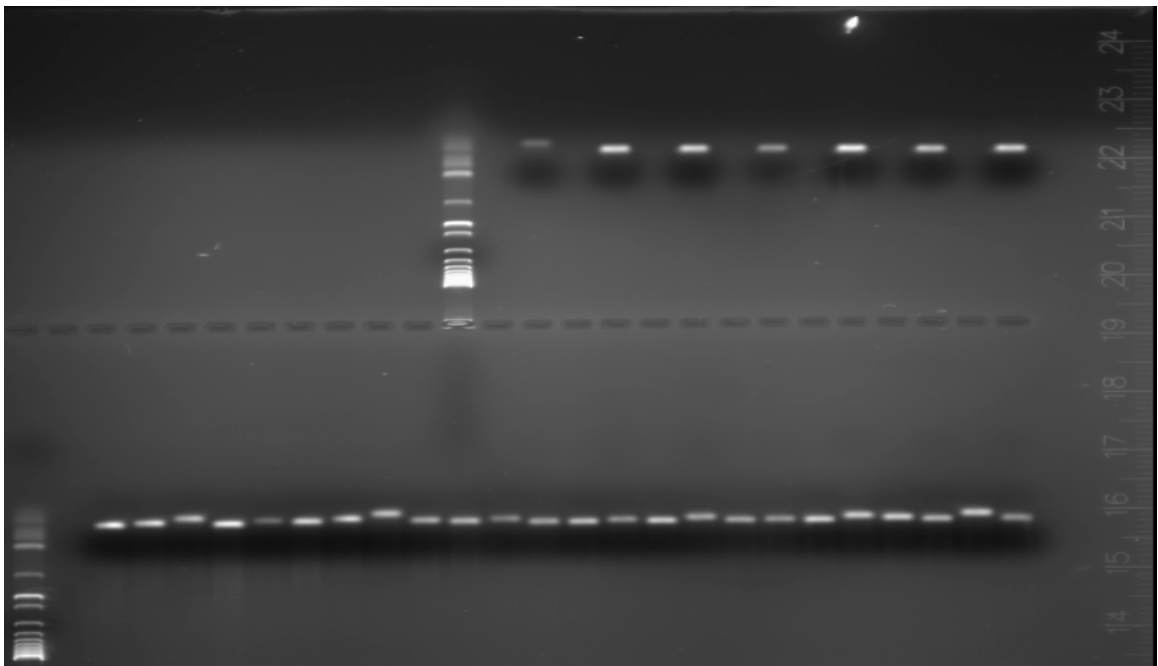
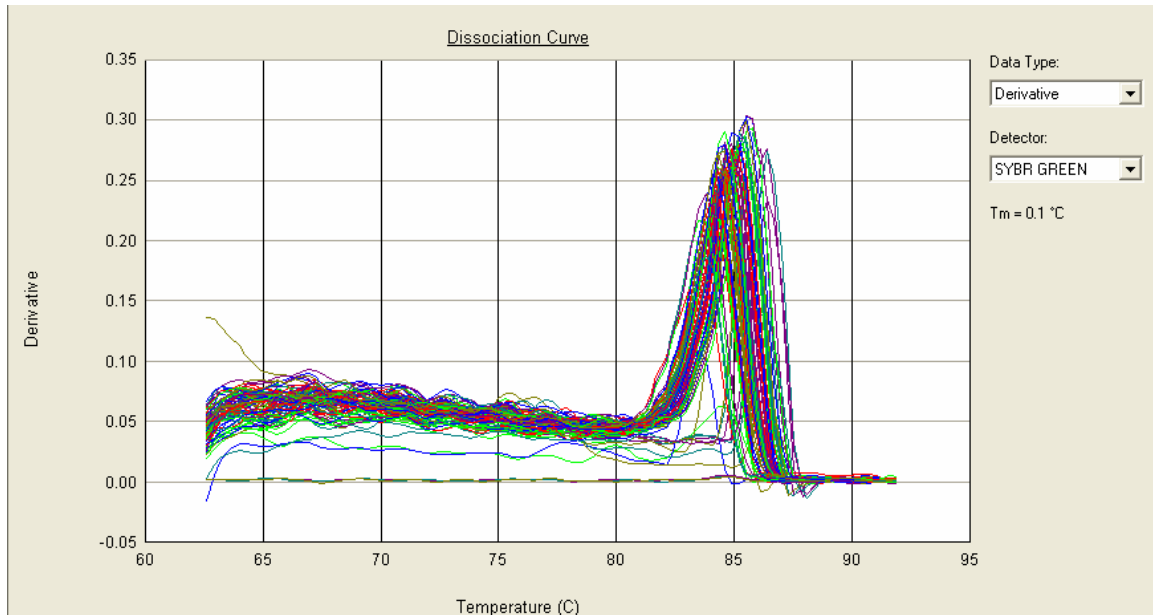


Figure 5: Dissociation Curve as seen in ABI Prism 7900 SDS software



In the above picture all the PCR amplicons were selected in the software interface, which is why multiple melting point curves are seen. Upon selecting individual PCR amplicons we can clearly see that each amplicon produces only a single product thereby adding information to the already existing gel electrophoresis image quality control data, stating the same.

4.) The calculation of DNase I sensitivity ratios is explained later in this discussion along with formulae.

The primary data analysis may be broken down into the following sub-phases: ^[1]

(Status: Test run with known DNase-I hypersensitive region primers (2) in progress and results from the first run are provided in the results section below.)

1.) Normalized fluorescence data were obtained using the ABI Prism 7900 SDS software and exported into Microsoft Excel worksheet for further analysis. The procedure for exporting this data is also mentioned below.

- 2.) An amplification curve and nth-order polynomial fit was obtained for each amplicon through the use of the ABI Prism 7900 SDS software.
- 3.) The C_t (cycle threshold) values were then extracted for each curve through the ABI Prism 7900 SDS software and exported into Microsoft Excel worksheet for further analysis.
- 4.) The amplification efficiency of the reference RHO (Rhodopsin) amplicon was obtained for every reaction plate using 5 replicates each of a standard dilution series of DNA amounts in the range: 0.3 ng, 3 ng, 30 ng, and 300 ng.
- 5.) We extract the 'clipped' data from the "delta Rn Vs. Cycle number" graph from the ABI 7900 HT SDS – Sequence Detection Systems software version 2.3 through the following procedure and then import it into the LinRegPCR^[5] software:

- 1.) Start the SDS 2.3 Software program.
- 2.) Click on File -> Open and then select the SDS data file to process.
- 3.) Click on the Analysis tab and click on Analyze.
- 4.) To export data to *.xls format :
 - a. Select the wells to export data from or you could export data from the entire plate.
 - b. Click on File -> Export, an Export window will open up.
 - c. In the Export drop down menu select data type "Clipped"
 - d. Select radio button option in the From: section to either "All Wells" or "Selected Wells".
 - e. Give a suitable file name and save the file in the Tab-delimited Text format *.txt
 - f. Click on the Export button.
 - g. Navigate to the saved *.txt file and open it.
 - h. In the opened text file click on Edit -> Select All, Copy.

- i. Open Excel and Paste the data. The Excel file will now contain Rn values per amplification cycle and Delta Rn values per cycle for every amplicon.
- j. Select only the Delta Rn values and paste them in a new excel worksheet, save the excel worksheet as a *.xls file and then open the LinRegPCR software.
- k. The excel file thus obtained can be exported into the LinRegPCR software and looks exactly like the snapshot taken shown below in Figure 6.

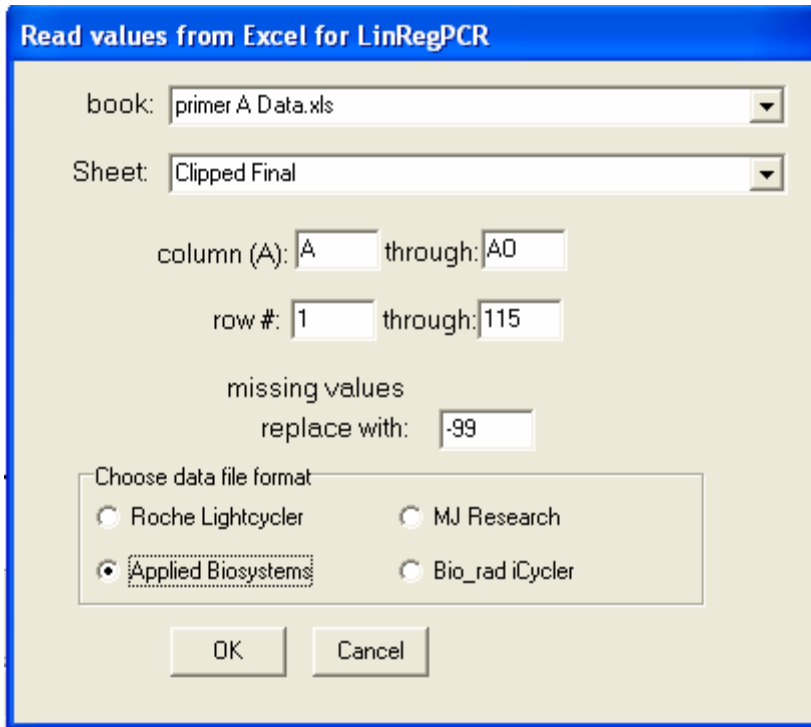
Figure 6: Snapshot of clipped data from delta Rn vs. cycle number graph

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	SDS 2.3	Clipped	1															
2	Well Nb	Reporter	Rn	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	19	Primer RHO		1.427079	1.484835	1.503058	1.50978	1.5246	1.539735	1.541034	1.539464	1.532475	1.524901	1.530278	1.528919	1.52655	1.526266	1.518773
4	21	Primer RHO		0.967426	1.013095	1.030083	1.041212	1.056159	1.059809	1.069991	1.071992	1.078259	1.082623	1.088465	1.083696	1.083815	1.085048	1.087739
5	23	Primer RHO		1.202619	1.249095	1.275963	1.288857	1.307442	1.311056	1.316837	1.322028	1.320073	1.318887	1.315461	1.314197	1.311368	1.307841	1.306413
6	43	Primer RHO		-0.03472	-0.0826	-0.0785	-0.12056	-0.07722	-9.80E-04	-0.10851	-0.15298	-0.11434	-0.14487	-0.10279	-0.2379	-0.09022	-0.13657	-0.14142
7	45	Primer RHO		0.104344	0.135494	0.10193	0.122838	0.030806	0.103943	0.128144	0.062843	0.057024	0.109369	0.048427	0.043184	0.062156	0.061694	0.04617
8	47	Primer RHO		-0.02577	0.00562	-0.01531	0.0797	0.100059	0.126096	0.031006	0.017792	-0.04142	-0.03682	-0.01011	0.037085	0.038192	0.037139	0.086635
9	67	Primer RHO		1.377789	1.418614	1.44251	1.450848	1.470632	1.484592	1.486322	1.486899	1.487277	1.477035	1.48711	1.48182	1.492228	1.480517	1.476093
10	69	Primer RHO		1.050789	1.085875	1.101795	1.114114	1.117731	1.124027	1.125466	1.133603	1.128024	1.133162	1.136726	1.131127	1.133161	1.131441	1.128367
11	71	Primer RHO		0.814257	0.841795	0.860894	0.867054	0.873483	0.878674	0.884532	0.88942	0.889091	0.893715	0.880339	0.864585	0.862474	0.865315	0.869858
12	91	Primer RHO		0.278965	0.301043	0.325179	0.303674	0.233381	0.31446	0.355902	0.309014	0.305341	0.394741	0.35085	0.212052	0.206931	0.224608	0.242663
13	93	Primer RHO		0.454217	0.377116	0.400044	0.33747	0.343376	0.414821	0.407005	0.377607	0.384973	0.450711	0.416042	0.330479	0.32052	0.397231	0.319476
14	95	Primer RHO		0.355447	0.373296	0.310458	0.26318	0.396234	0.303241	0.329825	0.320367	0.259904	0.42533	0.353014	0.470913	0.447367	0.423566	0.418658
15	115	Primer RHO		1.293966	1.312176	1.30332	1.322882	1.346529	1.353407	1.349716	1.356419	1.359385	1.360161	1.362588	1.370107	1.360738	1.366339	1.369217
16	117	Primer RHO		1.541787	1.597625	1.625554	1.648323	1.658514	1.669031	1.677054	1.684093	1.684078	1.689536	1.69092	1.69514	1.683591	1.696978	1.684751
17	119	Primer RHO		1.801299	1.84712	1.874929	1.890729	1.901202	1.90242	1.913419	1.902726	1.901227	1.89734	1.898178	1.900174	1.90311	1.896859	1.900297
18	139	Primer RHO		0.205135	0.182669	0.253413	0.208257	0.099405	0.251372	0.295564	0.248452	0.244358	0.242462	0.234731	0.188868	0.123594	0.113885	0.202374
19	141	Primer RHO		0.533958	0.555322	0.448848	0.556972	0.635122	0.430391	0.50765	0.543343	0.462672	0.483517	0.522814	0.527727	0.43073	0.605882	0.454228
20	143	Primer RHO		0.460251	0.488996	0.522306	0.467945	0.570804	0.511612	0.523706	0.465408	0.490559	0.546704	0.559525	0.488873	0.513445	0.573873	0.569631
21	163	Primer RHO		1.423867	1.456875	1.4763	1.490493	1.48924	1.502647	1.512055	1.51164	1.507163	1.510882	1.508079	1.511612	1.501416	1.535385	1.526804
22	165	Primer RHO		2.026347	2.092925	2.140709	2.154483	2.164988	2.178937	2.190836	2.19037	2.210376	2.189966	2.199062	2.190935	2.1964	2.198485	2.195323
23	167	Primer RHO		1.568512	1.623176	1.65067	1.663909	1.680858	1.695215	1.698958	1.703579	1.708028	1.708453	1.704145	1.718478	1.706529	1.707736	1.705758
24	187	Primer RHO		0.53274	0.61179	0.623369	0.523165	0.528928	0.580997	0.534272	0.522252	0.624878	0.498741	0.4878	0.606744	0.641897	0.567731	0.649557
25	189	Primer RHO		0.457501	0.450002	0.54168	0.560944	0.50357	0.429121	0.493622	0.380783	0.489866	0.525251	0.447303	0.49898	0.617365	0.456214	0.492615
26	191	Primer RHO		0.340678	0.464702	0.410393	0.425614	0.313427	0.385255	0.458169	0.454145	0.378134	0.352718	0.291527	0.406367	0.419379	0.414696	0.386748
27	1	PrimerA		0.420653	0.415936	0.417044	0.414848	0.410115	0.414747	0.410043	0.411015	0.405956	0.407821	0.403949	0.404137	0.397832	0.398926	0.397344
28	3	PrimerA		0.272108	0.27097	0.269148	0.267049	0.265232	0.261042	0.263707	0.260161	0.261978	0.261114	0.259052	0.259841	0.2571	0.255304	0.258411
29	5	PrimerA		0.529325	0.529876	0.526233	0.523852	0.523605	0.522887	0.523706	0.533418	0.535503	0.528509	0.531918	0.528898	0.526873	0.524902	0.521114
30	7	PrimerA		0.397122	0.39721	0.392672	0.394781	0.393685	0.391434	0.392806	0.392965	0.393687	0.392555	0.389197	0.392382	0.388504	0.3913	0.386375
31	9	PrimerA		0.262319	0.260611	0.257991	0.257967	0.255639	0.258865	0.255777	0.257562	0.253965	0.256486	0.25626	0.253702	0.255061	0.251755	0.249433
32	11	PrimerA		0.407887	0.402704	0.405271	0.405559	0.398157	0.405109	0.404419	0.397764	0.400018	0.404241	0.392068	0.393099	0.395151	0.389772	0.387494
33	13	PrimerA		0.294864	0.298753	0.291046	0.29172	0.292796	0.291854	0.28924	0.287204	0.286657	0.287453	0.283119	0.282492	0.280703	0.276604	0.278278
34	15	PrimerA		0.602777	0.598498	0.570413	0.577432	0.580068	0.574054	0.570846	0.57004	0.563457	0.561263	0.554963	0.550922	0.554072	0.55248	

5.) Reading Data into LinRegPCR software:

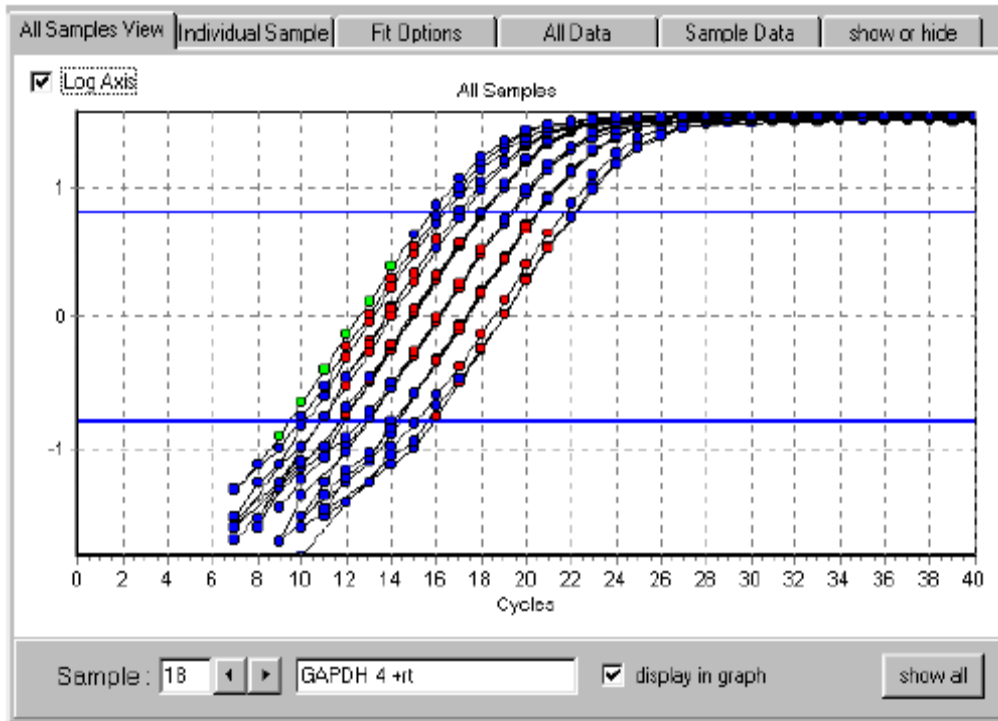
Once the data is exported into excel, do not close excel, open the LinRegPCR software and select the name of the excel worksheet from which data is to be imported in the Book: section of the first page of the software. Then specify the columns and rows to be imported from the excel worksheet and choose data file format as Applied Biosystems.

Figure 7: Export data from excel into LinRegPCR software



6.) Once the data is imported into LinRegPCR, this is how the 'All Sample View Screen' will look. This graph displays the data points of all imported samples. With the arrow keys you can browse through the samples. The 'Log Axis' check box at the top left hand corner of the graph toggles between the logarithmic and normal Y-axis.

Figure 8: ‘All Sample View Screen’ in LinRegPCR software



7.) Then click on the third tab: ‘Fit Options’ and select the default fit options and click on the ‘Fit All’ option which allows for automatically fitting all data points on every single curve according to the following criteria:

- a. The iterative algorithm of the software selects for at least 4 and maximum 6 data points
- b. with the highest R^2 value (note, R^2 value cannot be seen in this software during data analysis and this is done so as to not allow unfair manipulation of data.)
- c. And a slope, close to the maximum slope.

After automatic fitting of data one must look through every single curve to see if the 4 to 6 data points selected on every curve are within the linear portion of the curve and not the plateau phase of the S shaped amplification plot.

Here,

Intercept = Log (No)

i.e. No = $10^{\text{intercept}}$

[where No is the initial input amount of material.]

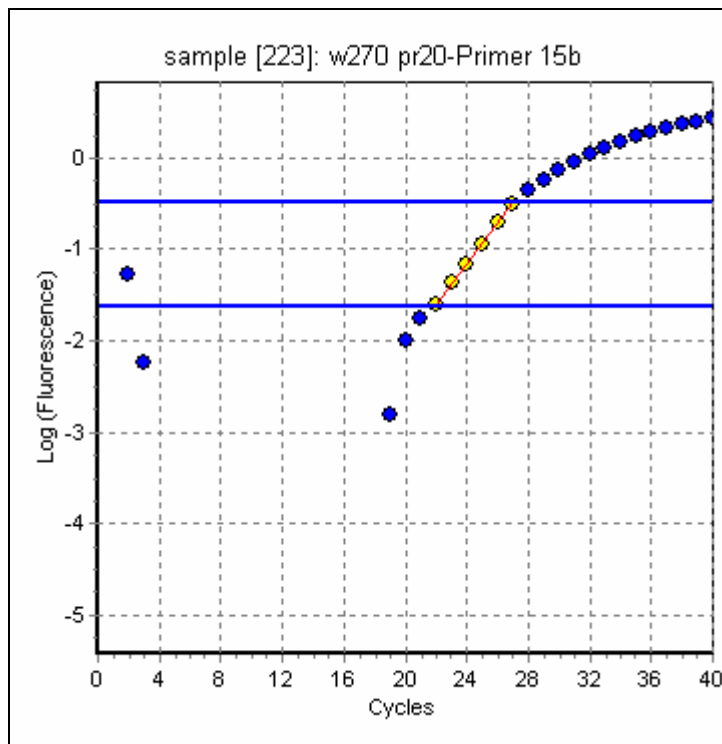
Slope = Log (E)

i.e. E = 10^{slope}

[where, E is the efficiency value of that sample at that given cycle threshold (Ct).]

- 8.) We derived the efficiency of each test amplicon from the amplification curve using the LinRegPCR software.

Figure 9: Selecting Data Points for each individual curve in LinRegPCR



- 9.) When all the adjustments to the data are complete, Choose File -> Save to Excel from the menu and choose the name of the excel workbook and worksheet to save your results into. The following picture shows how the final results sheet looks like. Following a three line header, one finds eight columns with sample number, sample name, lower and upper limit of window-of-linearity, number of included

points from that specific curve, starting concentration of the material No, PCR efficiency and correlation coefficient R squared.

- a. The output of the program gives the No values in fluorescence units the same as the output of the quantitative PCR machine is. To be able to convert to copy number or ng amount, we need a calibration curve. In our experiment though, we would be normalizing our results and giving them in the form of a ratio relative to the reference gene Rhodopsin and the DNase-I untreated samples, therefore we did not need a calibration curve as such ratios do not have a unit.

Figure 10: Final results output sheet with PCR efficiencies from LinRegPCR

Analysis of Real Time PCR data				version:7.5				
analysis date:10/1/2007								
fit option 2: n between 4 and 6 and best R								
sample	name	lower limit	upper limit	n included	No	PCR eff	R2	
1.000	w1 pr0.032062	0.070	0.500	5.000	2.86695E-08	1.633	0.999	
2.000	w2 pr0.031603	0.000	0.000	4.000	2.36015E-06	2.410	0.890	
3.000	w3 pr0.020179	0.010	0.000	9.000	0.016912114	0.812	0.742	
4.000	w4 pr-0.032059	0.010	0.120	4.000	4.81799E-15	2.353	0.999	
5.000	w5 pr0.00969	0.000	0.000	4.000	0.000353203	1.146	0.228	
6.000	w6 pr0.026444	0.010	0.190	4.000	3.21114E-13	2.335	0.997	
7.000	w7 pr0.000414	0.000	0.000	5.000	0.001092295	1.054	0.118	
8.000	w8 pr0.013775	0.000	0.000	5.000	0.000235117	1.188	0.480	
9.000	w9 pr0.076289	0.020	0.410	6.000	6.14127E-10	1.818	0.999	
10.000	w10 pr0.023188	0.010	0.470	6.000	7.534E-08	2.112	0.997	
11.000	w11 pr0.014803	0.000	0.010	5.000	0.000512664	1.186	0.733	
12.000	w12 pr0.011921	0.000	0.100	4.000	3.17236E-07	1.554	0.979	
13.000	w13 pr0.018601	0.020	0.480	6.000	1.87048E-07	1.849	1.000	
14.000	w14 pr-0.000882	0.040	0.280	4.000	5.23711E-08	1.961	0.999	
15.000	w15 pr0.393111	0.040	0.300	4.000	2.87522E-08	2.019	0.999	
16.000	w16 pr-0.084693	0.060	0.500	4.000	3.86264E-08	1.981	0.995	
17.000	w17 pr0.196027	0.020	0.210	4.000	2.64484E-09	2.203	1.000	
18.000	w18 pr-0.089781	0.140	0.560	4.000	7.65115E-06	1.565	0.999	
19.000	w19 pr0.022528	0.020	0.170	4.000	4.40541E-09	2.139	0.998	
20.000	w20 pr-0.004512	0.040	0.440	4.000	1.94535E-09	2.170	0.965	
21.000	w21 pr-0.001162	0.010	0.240	4.000	1.89643E-12	2.792	0.983	
22.000	w22 pr0.017121	0.020	0.310	5.000	3.50671E-07	1.919	0.998	
23.000	w23 pr0.0265	0.000	0.010	4.000	0.001697613	1.051	0.906	
24.000	w24 pr-0.053348	0.010	0.020	4.000	8.47769E-05	1.533	0.902	
25.000	w25 pr0.017639	0.000	0.010	4.000	0.000202257	1.242	0.789	
26.000	w26 pr0.022382	0.000	0.010	4.000	4.66835E-06	1.455	0.995	

10.) The general guidelines for selecting good data points from this final excel worksheet so as to perform efficiency correction on data points from all amplicons, are as follows: (personal communication, Dr. Jan Ruijter) ^[5]

- a. Samples with strongly deviating PCR efficiencies have to be considered unreliable and re-run if adequate data points aren't obtained from a given

PCR plate. The PCR efficiency variation permissible per amplicon should be within the median efficiency for that amplicon's range of efficiency values plus or minus 0.05.

- b. As per the author's ^[5] suggestion, samples with correlation coefficients below R=0.999 have to be considered suspicious because the points that are used to determine the efficiency (E) from must be on a very straight line. The difference between a line with R=0.999 and R=0.995 is not visible on most computer screens because the line is drawn in pixels. So one can only trust the R value, not what one sees. However, I have not used this criterion too strictly as we have found no supporting data requiring implementation of this stringent criterion.

11.) We then calculate the copy number differences arising due to DNase-I digestion, using the comparative C_t method using the expression given below.

12.) Efficiency corrected C_t values would then be used to compute a relative copy number ratio by applying the formula $2^{-\Delta\Delta C_t}$ ^[4] OR
 $2^{-[\text{DNase-I treated (target - reference RHO)} - \text{untreated (target - reference RHO)}]}$

Relative DNase-I sensitivity ratios thus obtained, having a value of less than one, would be indicative of relative copy loss resulting from digestion of that specific region by DNase-I.

The $2^{-\Delta\Delta C_t}$ method formula used in calculations is as follows ^[6,7,8,12]:

$$\begin{aligned} \frac{X_{N,q}}{RT} &= \frac{X_T}{R_0 \times (ER)^{C_{T,R}}} = \frac{\text{DNase-I treated target}}{\text{RHO treated}} \\ &= \frac{\text{Efficiency of primer A} [(10-80) \text{ DNase I conc}]^{C_{T,A} (10-80)}}{\text{Efficiency of RHO} [(10-80) \text{ DNase I conc}]^{C_{T,RHO} (10-80)}} \end{aligned}$$

= Normalized Target (Test)

$$\text{Similarly, } XN,Cb = \frac{\text{Untreated target}}{\text{Untreated RHO}} = \frac{\text{Efficiency of primerA (0 DNase I conc)}^{Ct A (0)}}{\text{Efficiency of RHO (0 DNase I conc)}^{Ct RHO (0)}}$$

= Normalized Target (Calibrator)

$$\text{And, } \frac{XN,q}{XN,Cb} = \frac{\text{Normalized Target (Test)}}{\text{Normalized Target (Calibrator)}} = 2^{-\Delta\Delta Ct}$$

Where,

XT = Number of Target Molecules at cycle threshold CT,X

RT = Number of Reference Molecules at cycle threshold CT,R

XN, q = Ratio of DNase-I treated

XN, Cb = Ratio of DNase-I untreated (calibrator)

X0 = Initial Number of Target molecules

R0 = Initial Number of Reference molecules

EX = Efficiency of Target at CT, X

ER = Efficiency of Reference at CT,R

Note: All the Ct values used in the formulae above are statistically corrected Ct values

The real-time PCR efficiency value should be close to 2.0 because at every stage in the linear part of the curve the amount of material (PCR product) is exactly doubled.

Since, the initial number of molecules i.e. amount is 30 ng/uL, the X0 and R0 can be eliminated from the above equation. And, the high expression cell lines for Kibra can be compared with the low expresser cell lines.

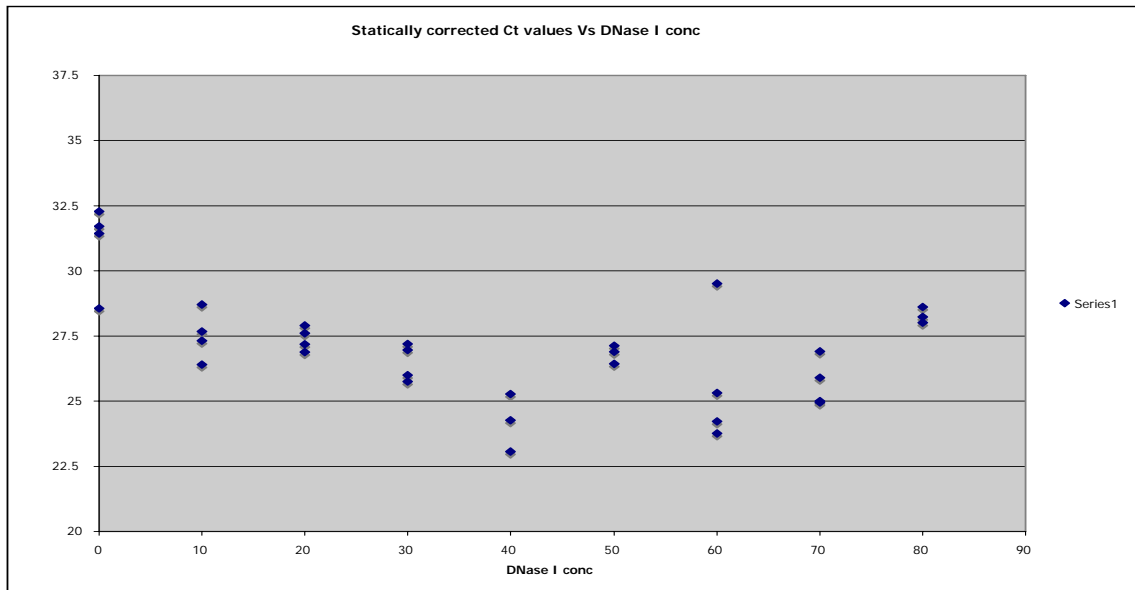
Results

Two amplicons A and B were designed across known DNase-I hypersensitive sites, which could be used as positive controls during the experiment. The data currently available is for the high expresser cell line MDA-MB-436 (details of which are given in the Appendix). We tested primer A within this cell line so as arrive at a single specific concentration of DNase-I that would help detect a hypersensitive site. We have the following number of good data points after statistical corrections to the data: (The results for the same are included in tabular form in Appendix-III)

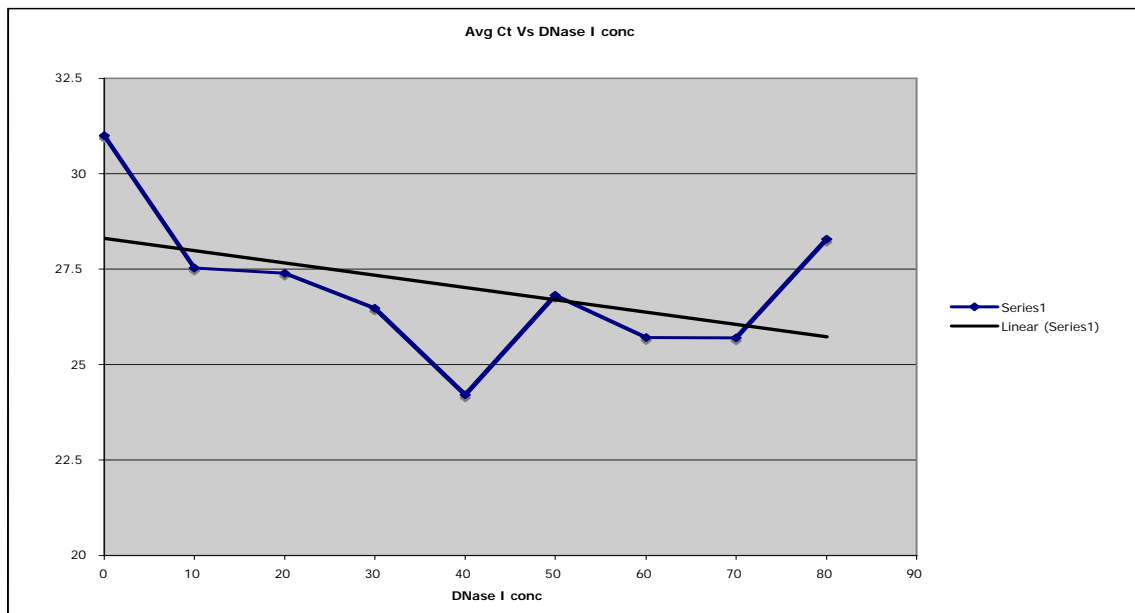
DNase-I Conc.	No. of Good Data Points
0	4
10	4
20	4
30	4
40	3
50	3
60	4
70	4
80	3
Rho 0	3
Rho 10	2
Rho 20	1

When we did 8 replicates on a 386 well plate, we got (3-4) good data points from the experiment. So when doing the next experiment, we would need to start with doing 8 replicates to get at least 3-4 good replicate data points. The layout for the 386 well plate has been provided in Appendix-II of this report.

After statistical corrections and eliminating data points whose R values were not close to 0.995 and above and whose efficiency values were not close to 1.8-2.0 range, the graph looks as follows:



When we take an average of these good data point values the graph we get is as follows:



As can be seen from this graph above, there is a drop in the Ct values at DNase-I concentration of 40 U/ml. We will continue to extract more data points at this sensitive concentration so as to verify this result and also test the 2nd positive control amplicon B (known hypersensitive region).

If the same results show up on amplicon B, we could use this specific DNase-I concentration to test the remaining 30 amplicons designed across chr5:167,769,118-167,801,709 in the Kibra gene so as to elucidate the locus control regions present.

Future Work – Recommendations

We continue to use the quantitative PCR method to elucidate the complete DNase-I hypersensitive profile at the Kibra gene locus, but doing so in multiple cell lines from different tissue samples is cumbersome. Testing in multiple cancer cell lines would get easier using a microarray approach. In fact up to eight different cells lines may be tested on a single microarray slide with replicates.

The following is a short comparison between the quantitative PCR approach using the $2^{-\Delta\Delta Ct}$ method and the microarray technology: ^[1, 2]

The main difference in both methodologies in terms of experimental setup is that the QT-PCR technology uses DNase-I treated versus non-treated cells and signal to noise ratios, whereas the microarray methodology would use the ratio between DNase-I treated nuclei -> extracted DNA (chromatin specific fragments) versus DNase-I treated naked genomic DNA fragments.

The cells would need to be synchronized in the G1 phase for the microarray methodology, whereas for QT-PCR via the $2^{-\Delta\Delta Ct}$ method, the cells are not synchronized. The reason why it would be a good idea to synchronize the cell growth phase would be because the exposure of some non-coding functional elements on the chromosome would be different in different cell growth phases i.e. as compared from G1 to M phase.

There is evidence that the microarray method allows for a great deal of reproducibility ^[2] i.e. if you repeat the same experiment even with a different batch of cultured cells the difference between the information collected out of the two experiments was shown to be about 6%. The accuracy of both methods is commendable when compared against previously existing Southern blot data as verified by experiments in referenced articles 1 and 2.

There are no southern blot experiments done on the Kibra that give information about

where the DNase-I hypersensitive sites might be within the 32.4 kb region and therefore placing positive controls on the microarray slides would be essential. These would mainly be known DNase-I hypersensitive regions which have been previously explored through other genetic experiments. A list of some of these positive control LCR's is also provided in Appendix-IV of this report.

The computational part for the microarray experiment would be far more intense as compared to the QT-PCR method as we would need to analyze patterns in consecutive microarray spots along with differences in color intensities so as to ascertain presence of hypersensitive regions. We would use the Axon GenePix Pro program to extract the signal intensities from the microarray scanned images. For the secondary data analysis, we would use R code.

One of the few common advantages of using the microarray method is that it provides more data through increased replicates. Experimental and personal error like handling etc. is also reduced in the microarray method, because the total number of steps involved in the bench-work for a microarray experiment is far less as compared to the QT-PCR approach, and are fairly standardized and designed to give a specific output.

The ENCODE ^[19] (Encyclopedia of DNA Elements) project would also lead to providing direction in the form of developing many more high throughput approaches for detecting locus control regions within the entire human genome, thereby leading to a better understanding of the regulatory mechanism of several genes involved in disease processes.

References

1. Dorschner et. al. (Dec 2004). **High throughput localization of functional elements by Quantitative Chromatin Profiling**, *Nature Methods*, Vol 1. No. 3.
2. Sabo et. al. (July 2006). **Genome scale mapping of DNase-I sensitivity in vivo using tiling DNA Microarrays**, *Nature Methods*, Vol 3. No. 7.
3. Marisa Wong and Juan Medrano (July 2005) **Real-time PCR for mRNA quantitation**, *BioTechniques* Vol. 39, No. 1
4. Livak and Schmittgen (2001) **Analysis of Relative Gene Expression Data Using Real Time Quantitative PCR and the $2^{-\Delta\Delta C_t}$ Method**, *Methods* 25, (Pg. 402-408).
5. Jan M. Ruijter et. al. (2003) **Assumption-free analysis of quantitative real-time polymerase chain reaction data**. *Neuroscience Letters*. 339, (Pg. 62 – 66).
6. Applied Biosystems (1997): **ABI User Bulletin #2 – ABI PRISM 7700 Sequence Detection System**.
7. Applied Biosystems: **Guide to performing relative quantitation of gene expression using real time quantitative PCR**
8. Michael W. Pfaffl **A-Z of Quantitative PCR, Chapter 3 – Quantification strategies in real-time PCR**, *International University Line (IUL), La Jolla, CA, USA*.
9. Qiliang Li, Kenneth R. Peterson, Xiangdong Fang, and George Stamatoyannopoulos (November 2002) **Locus Control Regions**, *Blood*, Volume 100, Number 9, (Pg. 3077 - 3086).
10. Michael McArthur, Shawn Gerum and George Stamatoyannopoulos (2001) **Quantification of DNaseI-sensitivity by Real-time PCR: Quantitative Analysis of DNaseI-hypersensitivity of the Mouse Beta-Globin LCR**, *Journal of Molecular Biology*, Vol. 313 – (Pg. 27-34).
11. Michael W. Pfaffl (2001) **A new mathematical model for relative quantification in real-time RT-PCR**, *Nucleic Acids Research*, Vol. 29. No.9 (Pg. 2002-2007)
12. Roche Applied Science – Technical Note No. LC 13/2001 - **Relative Quantification – LightCycler**.

13. Stuart N. Peirson, Jason N. Butler and Russell G. Foster (2003) **Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis**, *Nucleic Acids Research*, Vol. 31, No. 14 e 73.
14. Ronald H. Lekanne Deprez et. al. (2002) **Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions.**, *Analytical Biochemistry*, Vol. 307, (Pg. 63-69).
15. Joshua S. Yuan et. al. (2006) **Statistical analysis of real-time PCR data – BMC Bioinformatics**, 7:85
16. Ralf Gilsbach et. al. (February 2006) **Comparison of in vitro and in vivo reference genes for internal standardization of real-time PCR data – BioTechniques Vol. 40 No.2:** (Pg. 173-177).
17. Marc Reitman et. al. (July 1993) **An Enhancer/ Locus Control Region is not sufficient to open chromatin**, *Molecular and Cellular Biology*, Vol.13, No.7, (Pg. 3990-3998).
18. **The ENCODE (ENCyclopedia Of DNA Elements) Project – Science**; October 2004, Vol. 306, Issue 5696, (Pg.636-640).
19. Duncan Sproul, Nick Gilbert and Wendy A. Bickmore (October 2005). **The Role of Chromatin Structure in Regulating the Expression of Clustered Genes**, *Nature Reviews – Genetics*, Vol. 6 (Pg. 775-781).
20. Gary Felsenfeld & Mark Groudine (January 2003) **Controlling the Double Helix**, *Nature Vol. 421*, (Pg 448 – 453).
21. Marshall Elliot (October 2004) **Getting the Noise out of Gene Arrays**, *Science*, Vol. 306, Issue 5696, (Pg 630-631).
22. Ren Bing et. al. (December 2000) **Genome-Wide Location and Function of DNA Binding Proteins**, *Science*, Vol. 290, Issue 5500, (Pg 2306-2310).
23. Gary Felsenfeld (January 1992) **Chromatin as an essential part of the transcriptional mechanism**, *Nature*, Vol. 355, (Pg. 219-224).
24. Sarah C. R. Elgin (December 1988) **The Formation and Function of DNase I Hypersensitive Sites in the Process of Gene Activation**, *The Journal of Biological Chemistry*, Vol. 263, No. 36, (Pg. 19259-19262).
25. Adam G. West and Peter Fraser (2005) **Remote Control of Gene Transcription**, *Human Molecular Genetics*, Vol.14, Review Issue 1 (Pg. R101-R111).

26. David S. Gross and William T. Garrard (1988) **Nuclease Hypersensitive Sites in Chromatin**, *Annual Review of Biochemistry*, Vol. 57, (Pg. 159-197).
27. Peter J. Sabo, Richard Humbert, Michael Hawrylycz, James C. Wallace, Michael O. Dorschner, Michael McArthur, and John A. Stamatoyannopoulos (2004) **Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries**. PNAS 101:4537-4542.
28. Jan M. Ruijter and C. Ramakers (2003). **LinRegPCR: Analysis of real-time quantitative PCR data**. User Manual provided with software.
29. Michael W. Pfaffl, Graham W. Horgan , and Leo Dempfle (2002). **Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR**. Nucleic Acids Research 30: No. 9 e36.
30. Katrin Büther , Christian Plaas , Angelika Barnekow and Joachim Kremerskothen (2004). **KIBRA is a novel substrate for protein kinase C ζ** . *Biochemical and Biophysical Research Communications*. Volume 317, Issue 3, Pages 703-707.
31. Joachim Kremerskothen, Christian Plaasa, Katrin Büthera, Indra Fingera, Stefan Veltela, Theodoros Matanisa, Thomas Liedtkeb and Angelika Barnekowa. (2003) **Characterization of KIBRA, a novel WW domain-containing protein**. *Biochemical and Biophysical Research Communications*, Volume 300, Issue 4, Pages 862-867.
32. Andreas Papassotiropoulos et.al. **Common Kibra Alleles Are Associated with Human Memory Performance** (2006) *Science* 314 (5798), 475.
33. Suresh K. Rayala, et.al.(2006) **Essential Role of KIBRA in Co-activator Function of Dynein Light Chain 1 in Mammalian Cells**. *J. Biol. Chem.* 281: 19092-19099.
34. Peter J. Sabo et.al. (2004) **Discovery of functional noncoding elements by digital analysis of chromatin structure**. *Proc Natl Acad Sci U S A.* 2004 November 30; 101(48): 16837–16842.

Appendix - I

Cell Lines and Cell Culture Methods:

The four cell lines used for the experiment are:

1. IMR-90 Normal Lung Fibroblast [Low Expresser of Kibra]

[Ref: <http://www.atcc.org/common/catalog/wordSearch/results.cfm>]

Morphology: fibroblast

Species: human, Caucasian female 16 fetal weeks;

Tissue: lung, embryonic

Cytogenetics: normal human female; diploid; stable

2. T-47D Breast Cancer [High Expresser of Kibra]

[Ref: <http://www.biotech.ist.unige.it/cldb/cl4923.html>]

Morphology: epithelial-like

Species: human female 54 years old;

Tissue: breast; Tumor: carcinoma, ductal;

Derived from: T-47, originally derived from pleural effusion;

Cytogenetics: Hypotriploid human cell line. Y chromosome is absent.

3. U-87 MG Glioblastoma [Low Expresser of Kibra]

[Ref: <http://www.biotech.ist.unige.it/cldb/cl5280.html>]

Morphology: epithelial

Species: human, Caucasian female 44 years old;

Tissue: brain; Tumor: glioblastoma-astrocytoma;

Tumor stage: grade III

Cytogenetics: Hypodiploid human cell line.

4. MDA-MB-436 Breast Cancer [High Expresser of Kibra]

[Ref: <http://www.atcc.org/>]

Morphology: pleomorphic with multinucleated component cells

Species: human, Caucasian female 43 years old

Source Organ: mammary gland; breast

Disease: adenocarcinoma

Derived from metastatic site: pleural effusion

Cellular Products: tubulin; actin

Cytogenetics: The line is pleomorphic and most cells react intensely with anti-tubulin antibody as demonstrated by indirect immunofluorescence staining.

General Cell Protocols Followed throughout the experiment:

I. Storage of cells e.g. MDA 436:

1. Remove old media and rinse cells with 15 ml PBS to remove all FBS.
2. Add 2.7 ml trypsin solution and incubate cells 5-7 minutes (warm trypsin) at 37°C.
3. Tap the flask to loosen the cells and once they have loosened, add 15 ml warmed media (37°C) to the cell suspension. Collect suspended cells in 15 ml centrifuge tube and keep on ice.
4. Pellet at 100 G for 5 minutes. Remove the supernatant.
5. Re-suspend cells in media containing 10% DMSO
6. Label cryogenic vials with the cell line and date. Add 1.5 ml of the DMSO containing cell suspension to each vial and seal.
7. Place vials in -80°C freezer and keep overnight. The following day, after 24 hours, transfer cells to liquid nitrogen for permanent storage.

Two sets of MDA 436 cells were frozen from two different T-225 flasks. Labeled set 1 and set 2. Placed in liquid nitrogen tank, rack N.

II. Thawing of cell line e.g. IMR-90 (Normal lung fibroblast)

1. Remove cells from liquid nitrogen storage and quickly thaw in 37°C water bath by gently agitating vial.
2. As soon as ice crystals melt, pipette into a culture flask containing pre-warmed growth media.

Perform thawing and placed cells in 20 ml growth media. Incubated at 37°C.

III. Passage of cells e.g. MDA 436

1. Wash cells with PBS 2X. Remove PBS from cells each time.
2. Add 2.7 ml trypsin to flask. Let it sit at 37°C for 5-7 minutes.
3. Add 5 ml media to cells and transfer 1 ml of cell mixture into new T-75 flask containing 20 ml warmed at 37°C media after Trypsinizing cells and allowing them to set for 5-7 minutes at 37°C.

Passing of cells is done when cells reach 90% confluency.

Quantitative PCR – Protocol:

1. Synthesize cDNA with Invitrogen's SuperScript III using Oligo dT primers. For 1000 ng RNA starting material in a 21 uL cDNA synthesis reaction, dilute with 112.3 uL H₂O, which yields [7.5 ng/uL], use 2 uL (15 ng RNA) of this dilution for each subsequent qPCR. This will allow analysis of about 22 genes using 2 uL per reaction in triplicate.
2. Reverse transcribe the standard curve sample along-side the experimental samples. The standard curve should be composed of the RHO samples treated at various DNase-I concentrations (10-80 I/U).

3. Run samples in 9 replicates with standard curve samples at 30 ng, 3 ng, 0.3 ng, 30 pg, 3 pg RNA starting material.
4. PCR Thermal cycling conditions:
Stage 1: 95°C for 10 minutes
Stage 2: 95°C for 15 seconds and then 60°C for 1 minute (40 cycles)
Stage 3: Add a Dissociation/ Melting Curve Stage with the following default settings, 95°C for 15 seconds, 60°C for 15 seconds and 95°C for 15 seconds.

QT-PCR Protocol:

- 315 uL – SYBR green mix (2X)
- 1.9 uL – Forward Primer [100 uM]
- 1.9 uL – Reverse Primer [100 uM]
- 227.2 uL – Distilled Water
- 13 uL – of above mixture to each well
- + 2 uL – cDNA (15 ng/uL cDNA)
- 15 uL (TOTAL VOLUME of the reaction)

Protocol for DNase-I Digestion and DNA purification ^[1, 10]:

DNase-I digestions were performed according to the following protocol:

E.g. Cell Line: MDA 436 (Breast Cancer)

1. After cultivation cells are spun down, washed twice with 1X FBS and counted.
2. Add 2.7 ml trypsin solution warmed to 37°C. Let it sit for 5-7 min in incubator at 37°C
3. Add 10 ml media (warmed to 37°C) and transfer to 15 ml conical vial.
2-T-225's into two 15 ml conical vials (~16 ml volume)
4. Spin at 500 G for 5 minutes.
5. Wash cells 2 times with 10 ml PBS. Spin at 500 G for 5 minutes

6. Re-suspend in 9.5 ml Buffer A (cold) [Buffer A contains: 15 mM Tris-Cl, pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, pH 8.0, 0.5 mM EGTA, pH 8.0, 0.5 mM spermidine, 0.15 mM spermine]
7. To release nuclei, add 3 ml 2X NP-40 (0.08% in buffer A) drop wise to each tube. Incubate on ice for 10 minutes.
8. Nuclei were centrifuged at 1000 G for 10 minutes. Carefully remove supernatant and add 15 ml cold Buffer A. Repeat centrifugation.
9. Re-suspend both tubes in 1 ml total Buffer A by placing 800 μ l buffer A into one tube, re-suspending and transferring entire contents into the next tube.
10. 2X DNase I buffer: 120 mM CaCl₂ ; 1.5 NaCl
for 50 ml: 6 ml 1 M CaCl₂
15 ml 5 M NaCl
29 ml H₂O
11. Transfer 100 μ l of nuclei from step 9 into 9 tubes labeled 0 – 80 (stands for “units of DNase I”)
12. Add 100 μ l of 2X DNase I Buffer to each tube.
13. Add 0, 2, 4, 6, 8, 10, 12, 14 and 16 μ l of DNase I to each tube labeled 0 – 80.
Invert gently to mix. Incubate for 3 minutes at 37°C.
14. Stop the reaction with 16.6 μ l – 5 M EDTA. Flip gently.
15. Add 12 μ l RNase A (10 mg/ml final conc.). Flip gently to mix. Incubate at 37°C for 30 minutes.
16. Add 25 μ l Proteinase K (200 μ g/ml). Incubate at 55°C, overnight.
17. Extract with phenol/ chloroform two times. Add 1/10 volume sodium acetate (3M).
18. Add 2.5 volumes (550 μ l) 100% Ethanol. Store at -20 °C for 30 minutes.
19. Spin for 15 minutes, 12000 RPM to precipitate DNA. Wash 2X with 70% Ethanol (1 ml).
20. Let it air dry. Re-suspend in 20 μ l TE.
21. Store at 4 °C.

22. Samples were quantified in triplicate at a time for a total of 9 replicates using the Applied Biosystems ABI 7900 system.

Appendix - II

Excel worksheet with 386 well PCR plate layout attached with the Report.

Appendix – III

Results from MDA-MB-436 cell line for Test Primer A attached with the Report

Appendix – IV

Known Locus Control Regions to be used as positive controls for the microarray approach, alongwith EST related data attached with the Report.