Sindh Univ. Res. Jour. (Sci. Ser.) Vol.49(3) 517-524 (2017)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

A-stack probes do not show significant effects on Affymetrix GeneChip Data

G. N. MOJAI[++], F. N. MEMON, Z.U.A. KHUHRO, M. Y. KHUHAWAR*

Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Pakistan

Received 6th March 2017 and Revised 17th August 2017

Abstract: Besides normal binding of a guanine with a cytosine, a guanine can interact to another guanine through Hoogsteen hydrogen bonding. Such an interaction produces a tetrad with four guanines in which each guanine binds with two other guanines. A stack of tetrads of guanines can then form unusual four-stranded structures known as G-Quadruplex structure. It is reported that these types of structures are not good for the well-known microarray technology used for gene expression measurements. In particular, it is declared that Affymetrix GeneChips® produces incorrect gene expression measurements due to G-quadruplex structures. A gene is represented on GeneChip® by a group of small nucleotide sequences called probes and probes with continuous guanines showed abnormal behavior that has been associated with the formation of G-quadruplex structures. These probes are not correlated with other members of their groups but they are correlated with each others. It showed that such probes are not good measures of gene expression.

Since it is witnessed that an adenine can also bind to another adenine, it is therefore essential to find out its effects on Affymetrix GeneChip®. This paper presents the results of the examination of data of various GeneChip designs for the possible effects of adenine-adenine interaction. It is found that Affymetrix GeneChip® data does not show abnormal behavior due to the presence of probes with continuous adenines. Unlike guanine-guanine interaction, the GeneChip® is found unaffected of adenine-adenine binding.

Keywords: Microarray, Affymetrix, GeneChip®, Adenine-Adenine interaction.

1. INTRODUCTION

Microarray technology allows to perform the analysis of thousands of genes in a single experiment rapidly and efficiently. There are many commercial companies who manufacture microarrays; Affymetrix is one of the well-established manufacturers (Gautier 2004). Scientists are using different technologies for gene expression measurements. Affymetrix GeneChip® is one of the popular choices used for this purpose (Wu 2007). Although there are different platforms of Affymetrix, this paper focuses on Affymetrix 3'array.

A gene is presented on a 3' array by a group of 11-16 nucleotide sequences (Wu 2007, Upton 2008). Each sequence is known as a probe and is consisted of 25 nucleotide bases. The entire group of probes that represents a gene is called a probe set. The Affymetrix uses Photolithography technique to manufacture its GeneChips. Because of that, on a single thumbnail size array, millions of different probes can be synthesized. Since years, the GeneChip platform has proved that this technology is a reliable and robust system which enables scientists to discover many new therapies and revolutions to be made by the scientific community (Dennise, 2006). However, problems are found in GeneChip data. One of these problems is incorrect measurements of gene expression due to Guanine-Guanine interactions and hence it has been suggested to improve Affymetrix GeneChip® by more careful selection of probes in order to make it more reliable for gene expression measurement (Gautier 2004, Wu 2007, Upton 2008, Upton 2009, Memon 2009, Memon 2010a, Memon 2010b, Shanahan 2012).

The binding of guanine to cytosine and adenine to thymine usually occurs through the famous Watson-Crick interactions in double stranded DNA (Luis, 2010). However, in single-stranded DNA sequence, a guanine can interact to another guanine through a Hoogsteen hydrogen bond. Such an interaction produces a tetrad with four guanines in which each guanine binds with two other guanines at about 90 degrees (Memon, 2009). A stack of tetrads of guanines can then form unusual four-stranded structures, known as G-Quadruplex structures (Gautier 2004 & Evgenia 2012). Various chip designs for human are found effected by the formation of G-quadruplex structures (Memon 2010b). Similarly, various chip designs of other mammals are also found affected due to G-quadruplex structures (Memon, 2010b).

Since it is also observed by the scientists that adenine can also interacts with another adenine (Misako 1981, Ksenia 2010), it is obvious to think about the behavior of probes having continuous adenines. It is expected that adenine-adenine (A-A) interaction may cause similar kind of behavior as seen by guanine-guanine interactions. Hence, this study proposes the

[++]Corresponding Author: Email: rajperghulamnabi@gmail.com

*Institute of Advance Research Studies in Chemical Sciences, University of Sindh, Jamshoro, Pakistan

investigation of A-A interactions on GeneChip data for expected misbehaviour.

2. MATERIAL

A huge amount of GeneChip data of various living organisms is available freely for further analysis. This study initially focuses on Human GeneChip® data, however, data of two other organisms' chips are also analysed to verify the results produced by Human chip data. These two organisms are Rice (Plant) and Pseudomonas Aeruginosa (Bacteria). Hence, GeneChip data of three different kingdoms are tested for the effects of A-A interactions.

A number of different files are generated for a particular design of GeneChip such as probe sequence file, chip definition file, CEL file, etc. However, only two types of files are required for this study.

- **Probe Sequence data**: Probe sequence file of a GeneChip contains the information of probe sequences on that particular design of GeneChip. It includes Probe set ID, x and y positions of a probe to locate that probe on the chip, probe sequence and some other information provided by Affymetrix. These files are available at Affymetrix website (www.Affymatrix.com).
- **Experimental data**: These files are generated during a wet-lab experiment of the GeneChip in which target sequences are hybridized for gene expression measurements. The hybridization level of each probe is scanned and recorded as a numerical quantity in an electronic file known as a CEL file. These numerical quantities/intensity values of probes of interest will be used for analysis to examine the expected effects of A-A interactions. These files are available at NCBI GEO (Gene Expression Omnibus) repository (Barrett, 2005). Around 504 .CEL files of each organism are downloaded for analysis.

3. METHODOLOGY

The following steps describe the approach that has been adopted for the analysis of A-A interaction on GeneChip data. All the analysis is done by designing in-house tools for getting the results and achieving objective of this study.

1. Probe sequence file of a selected GeneChip design will be examined to filter out the probes having exactly four continuous adenines (A-run) in their sequences (called A-stack probes).

2. Categorize A-stack probes into groups according to the position of A-run within the sequences of A-stack probes. The possible position of A-run within a probe sequence could be P = 1, 2, 3, 22. Hence, 22 groups of A-stack probes will be produced. For instance, group 1 represents to all the A-stack probes in which A-run is at position one. This categorization of A-stack probes into groups will help to identify if there is any particular effect of position of A-run in probe sequences.

3. A-stack probes that belong to different genes/probe sets should not be correlated with each other if GeneChips are safe from A-A interaction. Otherwise, the correlation among these A-stack probes will prove that A-stack probes are also problematic for GeneChip like the case of G-stack probes reported in (Memon 2010a, Memon 2010b, Shanahan 2012). To verify this, the correlation among all the possible pairs of A-stack probes of two groups will be calculated that will be averaged then. Due to 22 groups of A-stack probes, a 22 by 22 matrix (M) is created in which each element represents the average correlation coefficient of two groups of A-stack probes; probes that are members of one specified group with probes that are members of another specified group. For instance, element of the matrix represents the average correlation value between A-stack probes in groups 4 and 10.

4. In the final phase, the matrix M will be used to draw a contour plot to demonstrate the overall correlation surface of the selected GeneChip.

4. RESULTS AND DISCUSSION

Human GeneChip (Kingdom: Animal)

Affymetrix has introduced various designs of Human GeneChip. If A-stack probes cause any problem on GeneChip data, it can be most prominent and seen easily on the chip having maximum number of A-stack probes. It is therefore, probes sequence files of all available Human GeneChip designs are used to identify the A-stack probes in them. **(Table 1)** is showing the fraction of A-stack probes in all these Human GeneChips.

Table 1: Statistics of annotated and A-stack probes in various Human GeneChip designs

GeneChip	Total no. of annotated probes	No of A-stack probes	%
HG_U133_Plus_2	604258	26778	4
HG_U133A	247966	9733	4
HG_U95A	201807	10190	5
HG_U95B	201862	12245	6
HG_U95C	201867	13524	7
HG_U95D	201858	13800	7
HG_U95E	201863	12747	6

It can easily be seen from Table 1 that among a list of Human GeneChip designs, HG-U95C and HG_U95D have about 7% probes with A-stack in them – the maximum fraction of A-stack probes in Human chip designs. As compare to HG-U95D, more data is available of HG_U95C in repository; it is therefore HG_U95C is selected for further analysis.

The A-stack probes of HG_U95C is then further classified into 22 groups according to the position of A-run within the sequences of these probes. **(Table 2)** is showing the number of A-stack probes in each group; group 22 has the largest number of A-stack probes.

These 22 groups of A-stack probes are used to calculate an average correlation among all possible pair of probes in two groups. As a result, a 22 x 22 correlation matrix is generated which is presented in **(Table 3).**

Table 2: Group-wise distribution of A-stack probes of HG_U95C.

Position of A-run	Number of A-stack probes	Percentage
1	187	1.38
2	136	1.00
3	686	5.07
4	628	4.64
5	596	4.41
6	605	4.47
7	511	3.78
8	502	3.71
9	438	3.24
10	599	4.43
11	504	3.73
12	591	4.37
13	683	5.05
14	567	4.19
15	627	4.64
16	569	4.21
17	712	5.26
18	779	5.76
19	867	6.41
20	885	6.54
21	864	6.39
22	988	7.31

The maximum value in the entire correlation matrix is 0.1 that is not a significant correlation. The Human GeneChip (HG-U95C) is showing poor correlation among A-stack probes. This shows that the probe level data is not affected by the presence of A-stack probes in HG-U95C.

Table 3: The correlation matrix is showing the correlation co-efficient between the probes of all possible pairs of groups in which A-stack is located at a particular position in Human GeneChip (HG_U95C).

Pos. of A-run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
3	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0
4	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
6	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
7	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
8	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
10	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
11	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
12	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
13	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
14	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
15	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
16	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
17	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
18	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
19	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
20	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0
21	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(Fig. 1) is showing the overall correlation surface of HG_U95C in the form of contour plot of correlation among the A-stack probes of all possible pair of groups.

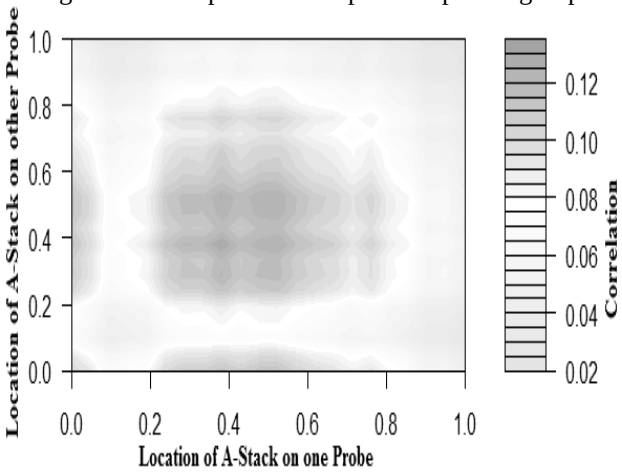

Fig. 1: Contour Plot showing Correlation surface of HG_U95C chip design.

Although the HG-U95C chip design shows higher fraction of A-stack probes as compared to the fraction of G-stack probes in various mammalian chip designs (Memon 2010b), the correlation surface of the HG-U95C chip is quite different. The plot in **(Fig.1)** is showing poor correlations among the A-stack probes in HG-U95 C. The highest correlation value appeared in this plot is 0.1.

Unlike various GeneChips of human which were found affected by the presence of G-stack probes, the contour plot of Human GeneChip does not show significant effects of A-stack probes at probe level data.

Pseudomonas Aeruginosa Genome Array (Kingdom: Bacteria)

Pseudomonas Aeruginose GeneChip is selected from the family of Bacteria to verify if bacteria GeneChips are also unaffected by A-A interaction. 1161 out of 77674 probes are found having A-stack in them. All the A-stack probes are divided into 22 groups **(Table 4).** These 22 groups of A-stack probes are then used to calculate an average correlation among all possible pair of probes in two groups. The 22 x 22 correlation matrix is presented in **(Table 5).**

The highest value in the whole correlation matrix is 0.15 that is again not a significant correlation. The Pseudomonas Aeruginose GeneChip is also showing poor correlation among A-stack probes like Human GeneChip (HG-U95C). This shows that the probe level data is not affected by A-stack probes in Bacteria chip as well.

Table 4: Group-wise distribution of A-stack probes of Pseudomonas Aeruginose GeneChip

Position of A-run	Number of A-stack probes	Percentage
1	0	0
2	1	0.08
3	70	6.02
4	76	6.54
5	54	4.65
6	46	3.96
7	44	3.78
8	38	3.27
9	36	3.10
10	73	6.28
11	42	3.61
12	59	5.08
13	87	7.49
14	47	4.04
15	44	3.78
16	55	4.73
17	40	3.44
18	55	4.73
19	73	6.28
20	70	6.02
21	64	5.51
22	87	7.49

Table 5: The correlation matrix is showing the correlation co-efficient between the probes of all possible pairs of groups in which A-stack is located at a particular position in Pseudomonas Aeruginosa Genome Array.

Pos. of A-run	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	-0	-0	0.1	0	0	0	0.1	-0	-0	-0	-0	-0	-0	-0.1	0	0.1	0	-0	0.1	-0
3	0	-0	0.1	0.1	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
4	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
5	0	0.1	0	0.1	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0
6	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
7	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
8	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
9	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
10	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0
11	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0
12	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
13	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
14	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
15	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0
16	0	-0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0
17	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
18	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
19	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
20	0	-0	0.1	0.1	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
21	0	0.1	0	0	0	0.1	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0
22	0	-0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0

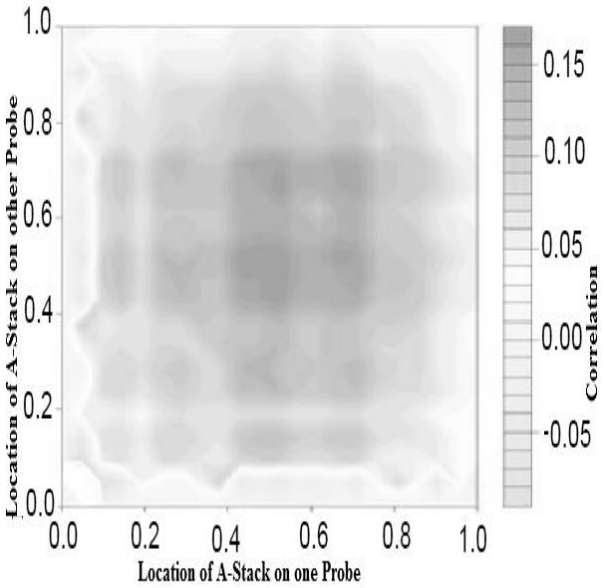

Fig. 2: Contour Plot showing Correlation surface of Pseudomonas Aeruginose GeneChip

A contour plot is presented in **(Fig. 2)** to show the overall correlation surface of Pseudomonas Aeruginose GeneChip. Although this chip design shows somehow similar proportion of A-stack probes as G-stack probes were found in various mammalian chip designs (Memon 2010), its correlation surface is different. The plot in Figure 2 is showing poor correlations among the A-stack probes in Pseudomonas Aeruginose GeneChip.

Rice GeneChip (Kingdom: Plant)

Finally, a Rice GeneChip data is taken from the plant family to verify the effects of A-A interactions. Table 6 is showing that 19795 out of 631066 probes contain A-stack in them. These A-stack probes are categories into 22 groups **(Table 6).** The average correlation of all possible pair of probes are calculated and presented in **(Table 7).**

The largest value in the whole correlation matrix is 0.2 that verifies a non-significant correlation among A-stack probes. The Rice GeneChip (Orysa Sativa) is showing poor correlation among A-stack probes **(Fig 3).** This shows that the probe level data is not affected by A-stack probes in Rice.

Althoug the Rice chip design show somehow similar fraction of A-stack probes as the fraction of

Table 6: Group-wise distribution of A-stack probes of Rice (Orysa Sativa).

Position of A-run	Number of A-stack probes	Percentage
1	2048	10.37
2	1470	7.44
3	916	4.64
4	702	3.55
5	633	3.20
6	659	3.33
7	587	2.97
8	628	3.18
9	756	3.83
10	813	4.12
11	804	4.07
12	773	3.91
13	811	4.10
14	773	3.91
15	565	2.86
16	695	3.52
17	690	3.49
18	635	3.21
19	733	3.71
20	833	4.22
21	1183	5.99
22	2088	10.58

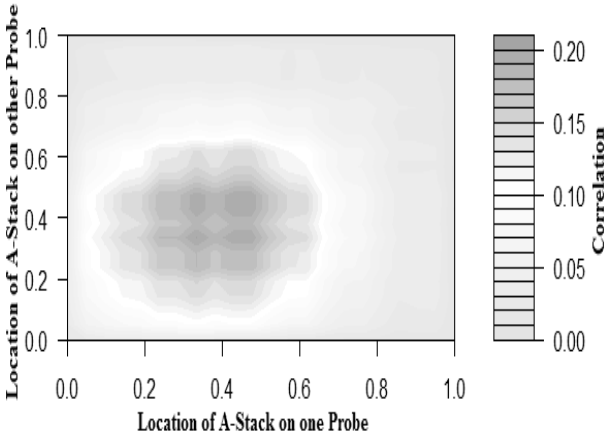

Fig.3: Contour Plot showing Correlation surface of Rice GeneChip.

G-stack probes in various mammalian chip designs (Memon 2010b) but unlike G-stack probes, the GeneChip data is unaffected by A-stack probes.

Table 7: The correlation matrix is showing the correlation co-efficient between the probes of all possible pairs of groups in which A-stack is located at a particular position in Rice GeneChip.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0
3	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0
4	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
5	0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
6	0	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
7	0	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
8	0	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0
9	0	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
10	0	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0
11	0	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0
12	0	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
13	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
14	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
15	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
16	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0
17	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0.1	0	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5. CONCLUSION

G-stack probes affect the probe level data as well as summarized data due to Guanine-Guanine interaction. This has been tested on various animal GeneChip data in general and mammalian GeneChip data in particular. Since Adenine can also bind with another Adenine, this study is designed to investigate similar type of effects. This paper has presented results of analysis of three different GeneChip data (Human, Pseudomonas Arguina and Rice to find out if the GeneChip data is also affected by the A-stack probe. These three Chip deigns showed poor correlation among A-stack probes. This shows that the probe level data is unaffected by A-stack probes and hence safe to be used for research and diagnostic purposes with the presence of A-stack probes.

REFERENCES:

Barrett T., T. O. Suzek, D. B. Troup (2005) "NCBI GEO: mining millions of expression profiles - Database and tools," *Nucleic Acids Research*, vol. 33, D562–D566,

Dennise D., D. Weiszhausz, J. Warrington, E. Y. Tanimoto, and C. G. Miyada (2006) "The Affymetrix GeneChipÒ Platform: An Overview", *Methods In Enzymology*, Vol. 410 doi: 10.1016/S0076-6879 (06) 10002-6

Evgenia N. N., F. L. Gottardo, H. M. Al-Hashimi (2012) "Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using.

Gautier, L., L. Cope, B. M. Bolstad, R.A. Irizarry (2004) "affy–analysis of Affymetrix GeneChip data at the probe level" *Bioinformatics*, 20(3):307–315.

Ksenia B. B., O Kostko, M Ahmed, Al Krylov (2010) "The effect of pi- stacking, H-bonding, and electrostatic interactions on the ionization energies of nucleic acid bases: adenine-adenine, thymine-thymine and adenine-thymine dimmers", *Physical Chemistry Chemical Physics*, 12(10), 2292-2307

Luis A. M., H.T. Lee, A. Garcia (2010) "Watson–Crick Base Pairs and Nucleic Acids Stability", Wiley online library

Misako AIDA and Chikayoshi NAGATA (1981) "Ab initio mo study on base stacking: adenine-adenine interaction in single-stranded polyadenylic acid (polyA)", Chemical Physics letters Vol. 86, Issue 1, 5 44-46

Memon F. N,, O. Sanchez-Graillet, J. G. Upton, and A. P. Harrison (2009) "Identifying the Impact of G-Quadruplexes on Affymetrix Exon Arrays using Computing". Journal of Integrative Bioinformatics.

Memon, F. N., A. M. Owen, O. Sanchez-Graillet, J. G. Upton, A. P. Harrison (2010a) "Identifying the impact of G-Quadruplexes on Affymetrix 3′ Arrays using Cloud Computing", Journal of Integrative Bioinformatics 7(2), 111-116.

Memon F., N. Graham. J. G. Upton, and A. P. Harrison (2010b) "A Comparative Study of the Impact of G-Stack Probes on Various Affymetrix GeneChips of Mammalia", Journal of Nucleic Acids, Vol. 20, Article ID 489736, 6Pp , doi:10.4061/2010/489736

NMR Relaxation Dispersion and Single-Atom Substitution", Journal of American Chemical Society DOI:10.1021/ ja2117816, 134, 3667−3670.

Upton G., W. Langdon, A. Harrison (2008) G-spots cause incorrect expression measurement in Affymetrix microarrays, BMC Genomics, 9, 613-616.

Upton Graham J. G., O. Sanchez-Graillet, J. Rowsell, J. M. Arteaga-Salas, N. S. Graham, M. A. Stalteri, F. N. Memon, S. T. May, A. P. Harrison (2009) On the causes of outliers in Affymetrix GeneChip data, 8(3), Briefings in Functional Genomics and Proteomics , 199-212.

Wu, C., H. .Zhao, K.. Baggerly, R. Carta, and L. Zhang, (2007) "Short oligonucleotide probes containing G-stacks display abnormal binding affinity on Affymetrix microarrays", Bioinformatics, 23, 2566-2572.