

## CHAPTER V

### RESULTS AND DISCUSSION

#### 5.1 MATERIAL SELECTION

Our group has a cohort of 300 young CRC patients with amplified DNA materials. This cohort has been used previously for CNVs analysis using SNP 6.0 and results obtained were robust. Due to the limitation of access to DNA materials, this cohort DNA material was used for our study using custom aCGH.

At the beginning of the study, aCGH analysis was completed on 41 amplified germline DNAs of CRC patients. CNVs found ranged between 39 and 1008 in a patient sample, giving an average of 320 CNVs observed per patient sample. Of these variations 9655 were DNA copy number losses. Within these losses, 401 of them with the size of 0.2 – 2.2 kb was found affecting miRNAs. On the other hand, 5387 DNA copy number gains were identified. These copy number gains were affecting 383 miRNAs with the size ranging from 0.2 – 2.6 kb. Many of the CNVs found were not able to be validated, leading to the notion of false positive. To direct this problem, we performed a direct match comparison of 14 samples using unamplified DNA materials. Results obtained from amplified and unamplified DNA material of the same patients were compared. CNVs that were unsuccessfully validated in amplified materials were not able to be trace back in unamplified materials results. Many of these CNVs that could not be found in unamplified materials were small in size and had few probes affected (5-15 probes). The total numbers of CNVs between the two material groups were

found to be significantly different (Supplemental 1). Consequently, only unamplified DNA materials results of the array were used from this point on.

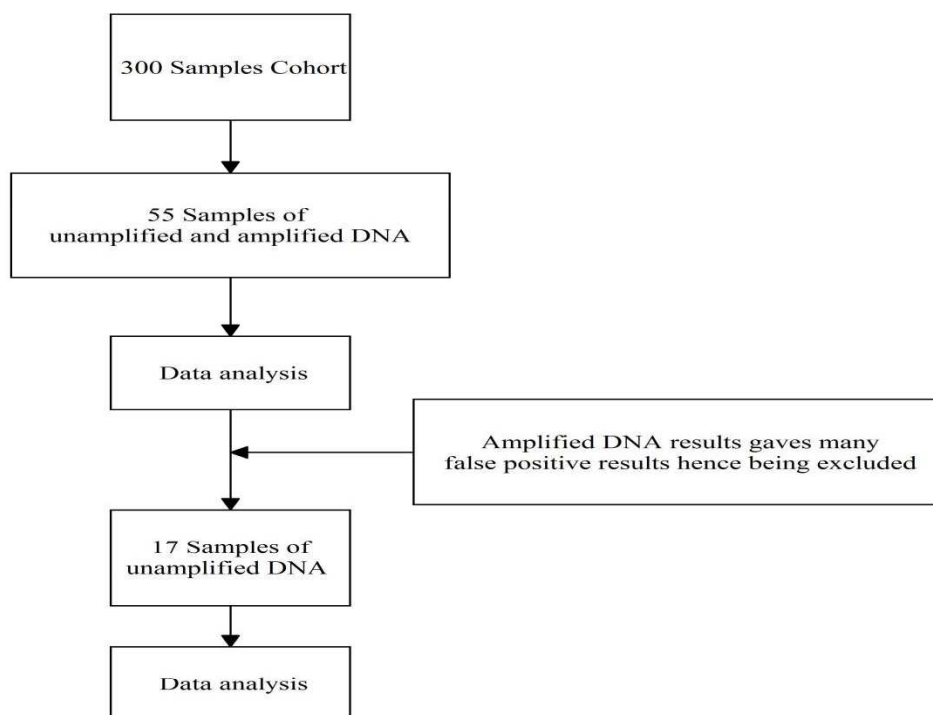


Figure 4. Flow chart of material selection. Originally 55 samples out of 300 cohort samples were used for the array using both amplified and unamplified DNA materials. After the data had been analysed, amplified DNA materials gave many false positive results, thus being excluded for the study. After the exclusion, as many as 17 unamplified DNA material samples were used and the data results being used throughout the study.

aCGH analysis was performed on 17 unamplified germline DNA materials of CRC patients. Segmental germline DNA copy number gains and losses were evident in all patients. It was revealed that 1018 copy number losses were found, in which 87 of those were affecting miRs. Majority of the copy number losses found affecting miRs had the size ranging between 0.2 - 1.2 kb (figure 5). Conversely, 745 copy number gains were found with 93 of those affecting miRs. Similar to the copy number losses, the sizes of copy number gains found on miRs were stretching between 0.2 – 2.2 kb (figure 6).

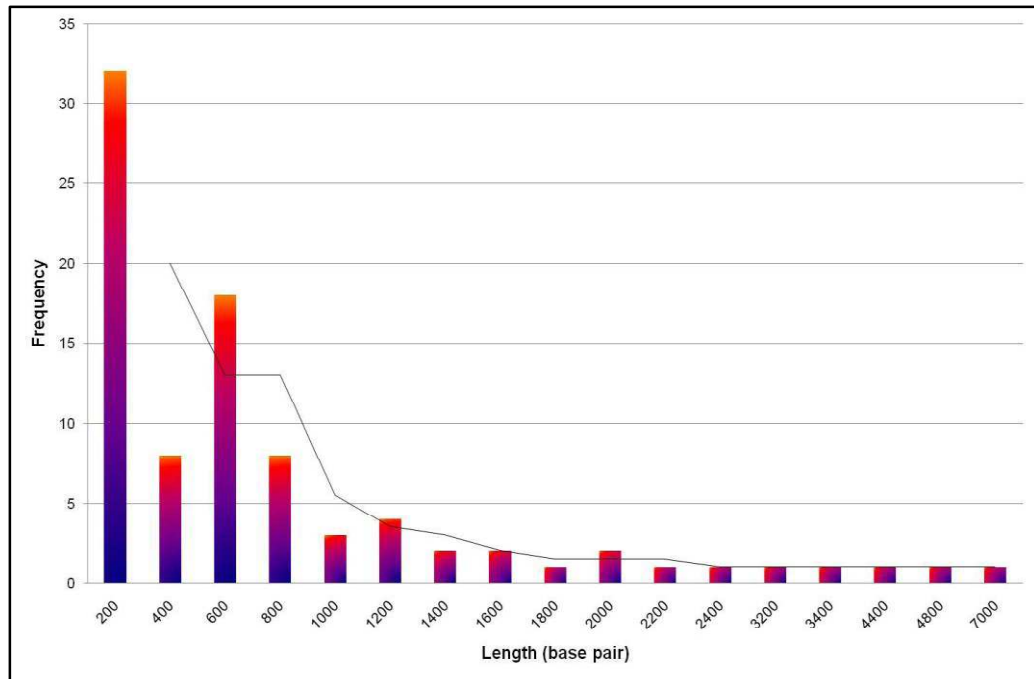


Figure 5. Size distribution of copy number deletion found affecting miRNAs. (N=17, Total # =87)

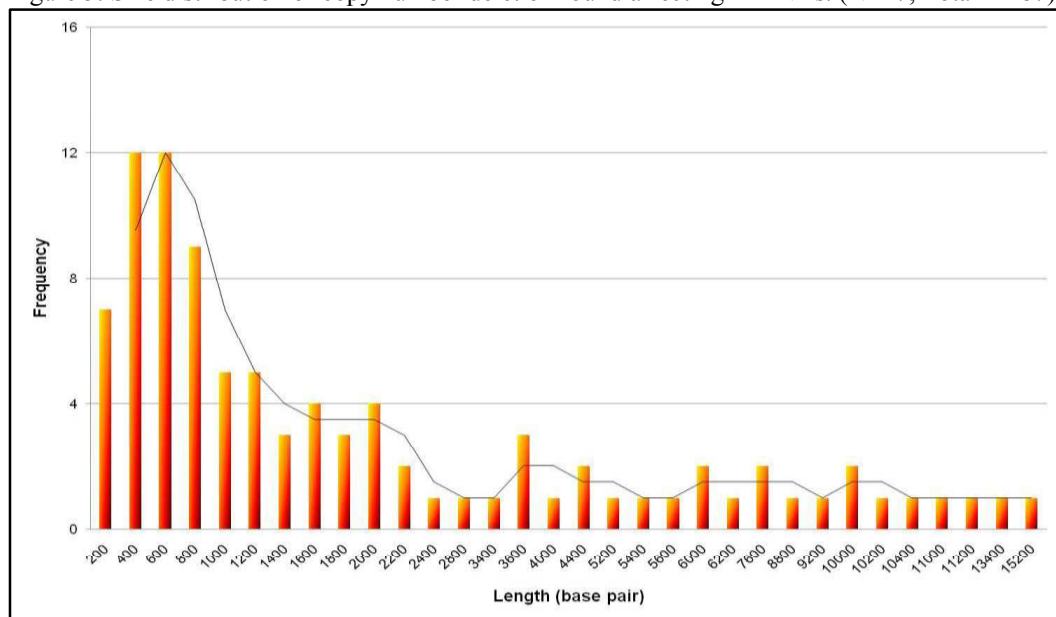


Figure 6. Size distribution of copy number duplication found affecting miRNAs. (N=17, Total # = 92)

## 5.2 GENERAL FINDING

Of the 17 samples that aCGH were carried out on; CNVs found ranged between 9 and 391. On average 103 CNVs were found on each patient sample. As

many as 1018 in copy number losses and 745 copy number gains were found using the unamplified DNA materials. The CNVs average of unamplified DNAs showed reduced number of lesions found on the array compared to amplified DNAs. Nonetheless, there were still many noises found on the array. Slight differences were found between the loss and gain size distribution in that most gains had a trend towards a bigger size of lesions compared to losses (figure 6). Next, CNVs found were divided into five groups: unique loss, recurrent loss, unique gain, recurrent gain, and recurrent in both loss and gain (table 1). Total miRNAs affected after divided into five groups were 82 miRNAs. Regardless, many of the CNVs found, even using unamplified DNAs were not validated therefore we assumed they were false positive. We discovered that the software provided by NimbleGen, made an individual calling of few lesions of the same kind (gains/losses) where those lesions could be joined into one calling of CNV.

### **5.3 RECURRENT COPY NUMBER VARIATIONS**

Of the five groups, CNVs with as many as 19 losses and 50 gains cases were found recurrently. These recurrent CNVs affected in total of 7 and 12 miRs for losses and gains. Recurrent CNVs losses affecting four miRs were unsuccessfully validated. Validations of other recurrent CNVs are still in progress. Furthermore we categorised other recurrent CNVs, that occur when both deletion and gain was found affecting one miR, into polymorphic CNVs. Majority of the CNVs found affecting miRs were indeed polymorphic CNVs. As many as 67 cases affecting 15 miRs were found recurrently affecting miRs through deletions and duplication.

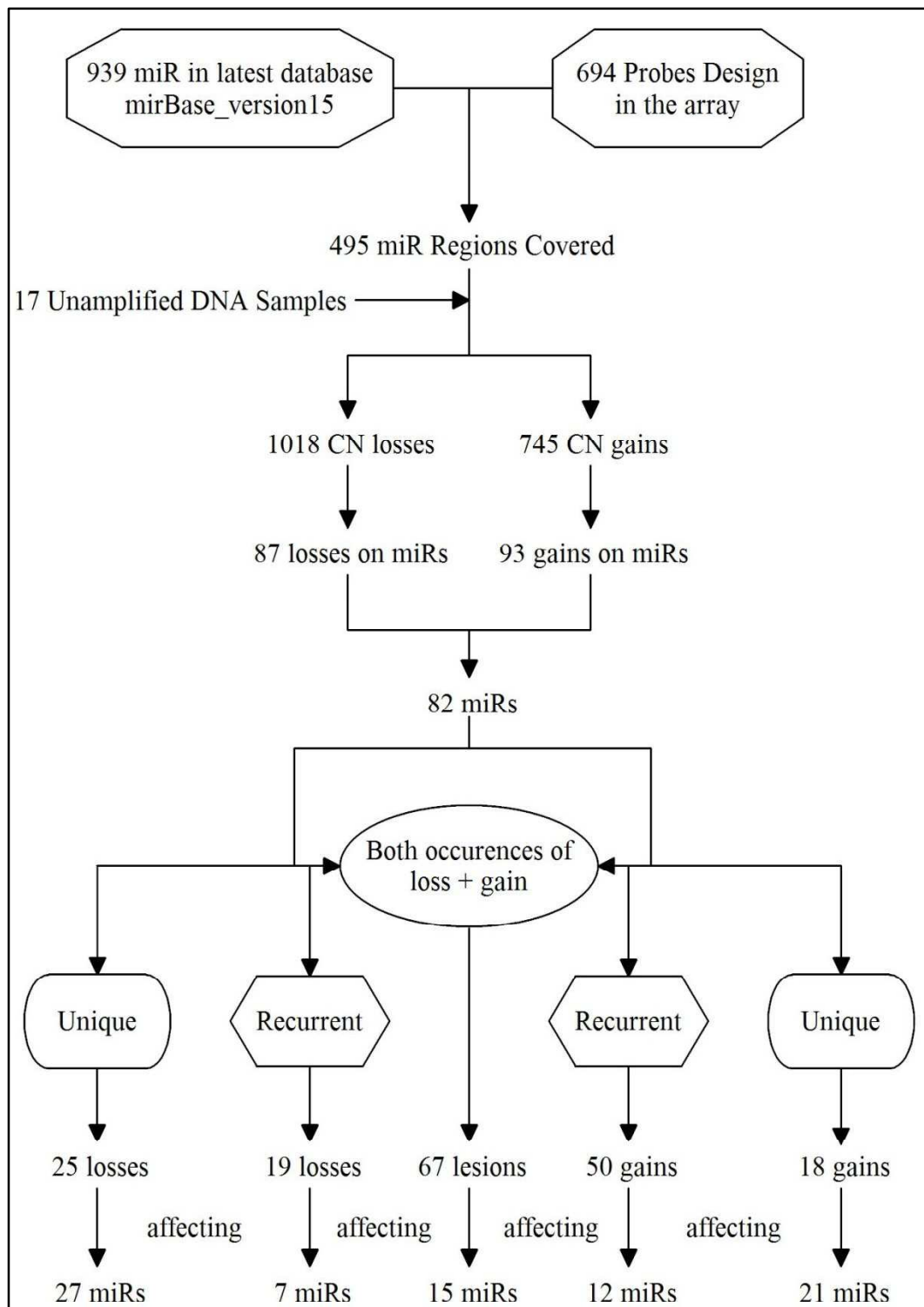


Figure 7. Flow Chart of unamplified DNA results. Probes on the array design is combined with the latest miR database, which give 495 miRs covered in the array. Through 17 samples of unamplified DNA materials, as many as 1018 deletions and 745 duplications were found, in which respectively 87 and 93 of those affecting miRs. After the CNVs affecting miR divided into five major groups, the totals of 82 miRs were affected by the CNVs.

We used an in house database of CNVs that had been performed in healthy cohort to check the results of the array. In the small cohort of patients, none of the recurrent CNVs was found on the database.

Table 1. Numbers of CNVs in five major groups. (N=17)

	Loss on miRs		Recurrences in both loss & gain	Gain on miRs	
	Unique	Recurrent		Recurrent	Unique
<b>Calls</b>	25	19	67	50	18
<b>Affected miRs</b>	27	7	15	12	21

#### 5.4 UNIQUE COPY NUMBER VARIATIONS

As many as 25 unique copy number losses and 18 copy number gains were found and affected 27 and 21 miRs respectively. One unique deletion found on the patient was validated although many were not. The CNV validated affected *miR-770* and this CNV was not found in any of the CNV database of healthy cohort. From this information, we gathered that the length of lesions combined with the probe mean value plays a role in determining the calling of real CNVs from false positive.

#### 5.5 CNVs VALIDATION

##### 5.5.1 Duplication

Previous experiments carried out in the lab [12] found a duplication affecting *miR-646*. The duplication found on the array stretches for 0.3 kb. A few MLPA primers were then designed around the duplication. The duplication was

able to be validated using the designed primers. Consequent screening in control showed no duplications were found on the 250 healthy control DNAs (figure 8). The focus of our study this time was on the deletions found on the array due to their less complexity to duplications.

### **5.5.2 Deletion**

Due to the less complex of deletion compare to duplication, we currently focused the validation of copy number losses found in the array. Quantitative PCR (qPCR) was used to validate one of the unique deletions. The deletion was located on chromosome 14 and it stretched for 1.5 kb (supplemental 3). Q-PCR primers were designed covering the deletion in miRNA. The deletion that affected miR-770 was able to be validated (figure 8). However, exact breakpoint of the deletion was unable to be mapped within the time frame. PCR primers were then designed following the result found using q-PCR. 94 healthy control DNAs using PCR primers were screened and no deletion was detected on these controls (figure 9).

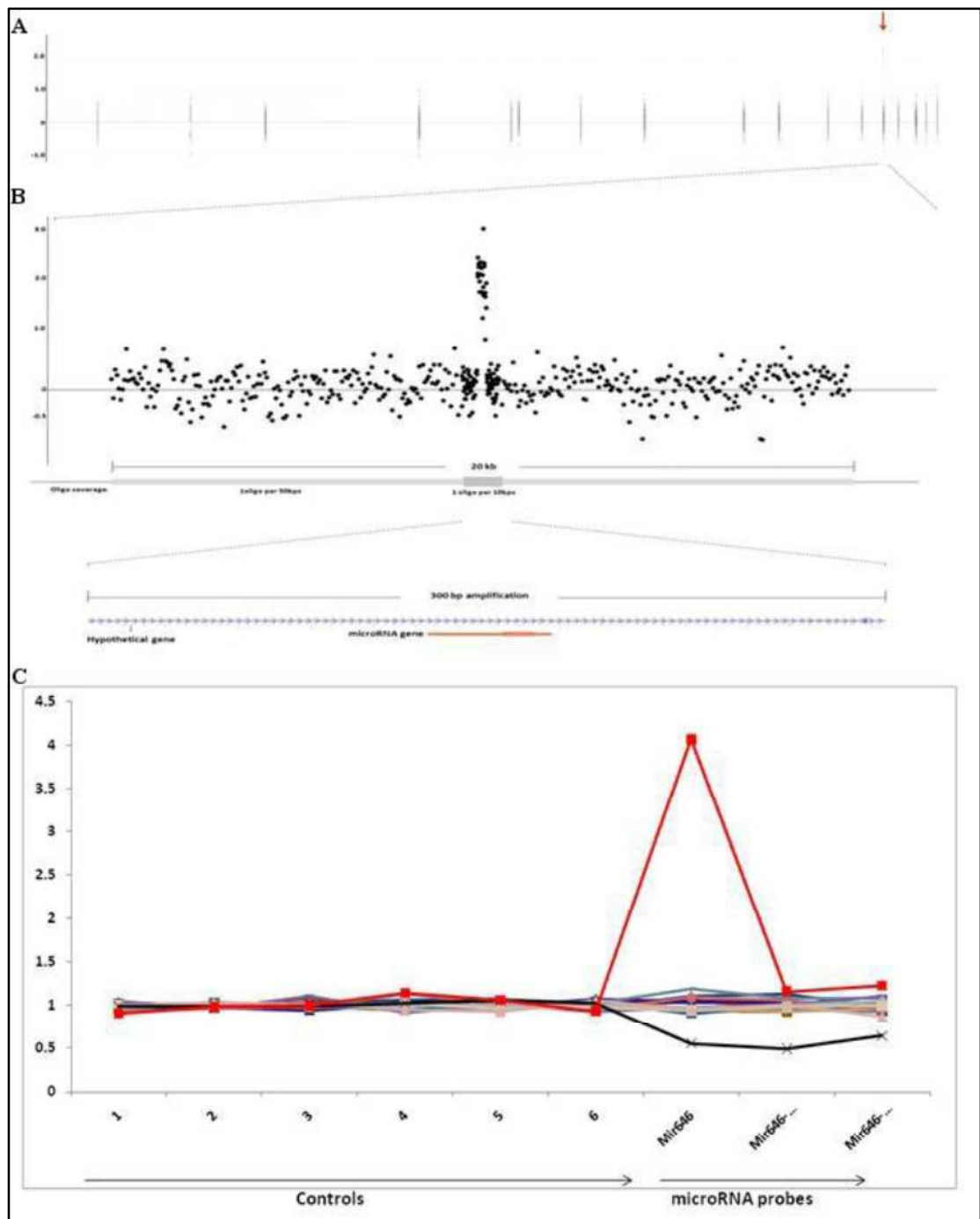


Figure 8. Validation of copy number duplication found affecting miR-646 on a patient. **A.** The chromosome representative of chromosome 14 obtained from signal map. Red arrow point to where the position of duplication was found **B.** A closer view of the duplication found on the patient. **C.** MLPA results of affected patient compare to 250 healthy controls, showing the probe covering miR-646 was duplicated in patient but not in controls.



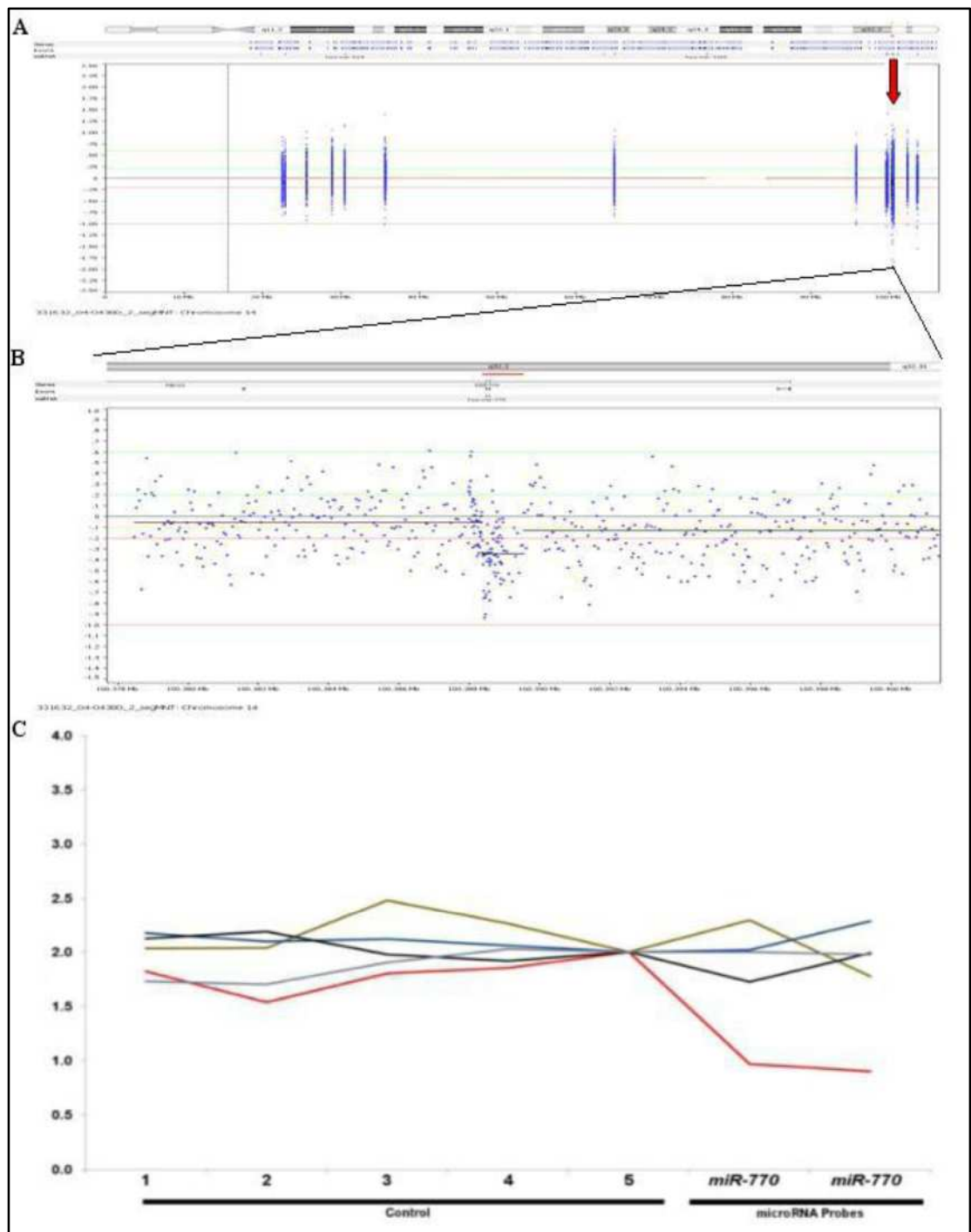


Figure 9. Validation of copy number loss found affecting miR-770. **A.** The nexus representation of chromosome 14 of the patient. The red arrow indicated the position of the deletion found. **B.** A closer view of the deletion affecting miR-770. **C.** Q-PCR result with 2 primers designed around miR-770, which shows heterozygous deletion compare to controls.

## 5.6 DISCUSSION

In this study copy number analysis affecting miRNAs genes using custom oligonucleotide array (oligo-aCGH) was proven to be a valuable technique to detect small size CNVs. These CNVs found were previously undetected on other array platform. To the best of our knowledge this array is the first to be able to detect small size CNVs targeting miRNA genes compare to other array platforms that has smaller resolution (~10-12 kb). Previous finding of duplication affecting miR-646 led to extensive study of CNVs affecting miRNA in familial CRC. In our study the patient cohort was selected carefully to include 17 patients who have CRC below the age of 40 years old. We found many CNVs affecting miRs that have small size and in addition many of these CNVs were not reported in any of the CNVs database known. Within this small patient cohort several unique CNVs affecting miRNAs in patients but not in controls were validated, namely miR-646 and miR-770. We have shown that these CNVs instead of being polymorphic CNVs may indeed play a role in familial CRC and its predisposition. Nevertheless, regardless of the fact that the array has high resolution, further study needs to be done to optimize the array and the results obtained.

Conversely, amplified DNA materials used at the start of the study gave a lot of noise on the array. In addition, several CNV candidates using PCR were validated and they were found to be false positive. A matched comparison of the amplified vs. unamplified samples revealed that many lesions detected by using DNA amplified samples materials were not found when unamplified DNA sample was used. A possible explanation for the observed phenomenon could be

that there is an improper amplification of the DNA samples causing the change on the DNA sequences. On the other hand we observed that unamplified material reduced drastically the false positive results and the background noise detected on the oligo-nucleotide specific array. Though the background noise was reduced, large number of CNVs detected per sample made it difficult to select the microRNAs affected for validation. Furthermore, the threshold settings for the array to distinguish real CNVs amongst the noise were not as straight forward as previously thought. We started validating CNVs of small size (0.05 kb) but were unsuccessful. To follow through, the CNVs affecting miRs with a bigger size (2 – 10 kb) were validated but didn't provide any results either. A change towards the validation approach was implemented through the use of lesion's length combined together with the probe median value. At the beginning validation of the CNVs with the lowest probe median value was done, however none of the CNVs were able to be validated. We noticed that CNVs with the lowest probe values had short lesion lengths found. Next, a selection was made of CNVs that have a low probe median value with length of >1 kb. This combination proved successful as a deletion affecting miR-770 was able to be validated using qPCR. Several primers were designed to validate the deletions and three primer pairs covering miR-770 showed a clear deletion. However, when PCR primers were designed a clear deletion on the patient was not able to be seen. The band of PCR product on the patient was observed and found to be lighter compared to healthy controls. Screenings of 94 controls were performed and none showed this problem encountered in the patient's DNA. Combining qPCR and PCR results in a

conclusion that even though qPCR showed a clear deletion, as to why it cannot be validated using PCR is due to the fact that the nature of the deletion on the patient may be more complex than originally expected. Instead of a deletion, this lesion can also represent a reverse translocation. Further study on this occurrence is currently still under way.

Our observation revealed that many CNVs indeed affecting miRNAs. Those CNVs have size ranging between 0.2 – 15 kb. For both duplication and deletions, most of the CNVs located had the length between 0.2 – 2 kb. These results confirmed the objective that the array was able to detect small size CNVs affecting miRNAs. These finding also showed that not only big size but also small size CNVs affect miRNA genes. A few years ago, small size CNVs that did not affect any genes were thought to be normal polymorphic CNVs without any possibility of being pathogenic. With the recent finding of miRNA genes many miRNAs have been associated with many diseases including CRC. Thus, small size CNVs might indeed play a role in diseases. We also evaluated the known polymorphic CNV covering miRNA genes using a compiled of available CNVs database (internet and results from other groups in the department). We found that 19 miRNAs from the miRNA database are indeed affected by polymorphic CNVs. With this information, we were not able to find these polymorphic miRNA CNVs back in our results. This can be contributed to the small cohort of patients that we performed the array with. Similarly, the unique lesions we found among the small cohort might mean that it's a recurrent lesion in a bigger cohort of patients. On the other hand if these lesions are not found using a bigger cohort, it will indicate

truly unique lesions. Therefore a bigger cohort of patients is needed to proceed further with the identification of unique or recurrent CNVs affecting miRNAs. Once we identified those CNVs we then have to screen large numbers of healthy controls to ensure the uniqueness of the CNVs.

Although most high-penetrant genes are frequently affected by point mutations, there are many reports of gross genomic rearrangements that may underlie cancer susceptibility. For examples, chromosomal rearrangement found on: 13q14 is associated with hereditary retinoblastoma [75], 11p13 is responsible for will's tumor [76], 5q21 is implicated in FAP [77], 10q23 is implicated in Cowden syndrome [78] and recently a micro deletion of *MSH2* gene in 2p21 is implicated in hereditary CRC [79]. Thus CNVs play a role in not only CRC but also other types of cancer predisposition. CNVs of smaller size affecting miRNA genes have been shown to exist. MiRNAs, a new player in the genomic world, have many functions that the dysregulation of it leads to the onset of many diseases, such as cancer. To the best of our knowledge, no study has been done on CNVs affecting miRNAs in germline DNA of CRC patients. This study will give us a new perspective of how CNVs affecting miRNAs which can be pathogenic or normal variations found in human. Either way this will give us more insight on CNVs on miRNA genes and its affect. Further study, such as validation and correlation with the pathological as well as clinical data will enable the use of the results to be used in clinical settings, such as counseling genetic, therapeutic, diagnostic marker, and prevention management.