8. Discussion

8.1 Oligoprobe selection

Although the use of DNA oligoprobes as short as six nucleotides has been reported (Drmanac et al. 1990), hybridisations with hexamer and heptamer DNA oligoprobes could not always yield consistent results due to the poor hybridisation stability. As a consequence, current OFP technology uses octamer (Clark et al. 2001), nonamer (Willse et al. 2004) or degenerated DNA decamer with an octamer core as hybridisation oligoprobes (Guerasimova et al. 1999 and 2001). Here, we show that with LNA modification, hybridisation of oligoprobes as short as heptamers on surface or even hexamers in solution, can be reliably used for DNA sequence characterisation.

8.1.1 Significance of oligoprobe length

Oligoprobes differ in hybridisation frequency and hybridisation quality. The selection of oligoprobes is an important feature of hybridisation experiments because it influences the clustering quality of fingerprints. The optimal oligoprobe length is related to the complexity of target DNA and its length. Informatical, biochemical, and technological factors determine the optimal lengths of the oligoprobes. Having targets immobilised onto the surface, oligoprobe lengths should be of 6-10 nucleotides (Drmanac et al. 1989). Ideally, 6- or 5-mer oligoprobes should be used. Short oligoprobes, however, are less specific and less accurate, which results in high signal noise and impeded data analysis. Although hybridisation of DNA oligoprobes as short as 6-mer (Drmanac et al. 1990) or 5-mer (Pe'er et al. 2003) were described, oligoprobe length is usually limited to 8–10 nucleotides due to the low stability of very short duplexes. In this study we observed rapid and cost-effective hybridisation-based characterisation using heptamer and even hexamer oligoprobes.

8.1.1.1 Shorter oligoprobes increase the hybridisation probability

The rationale of resequencing replies on possibility to extract similar amounts of sequence information from hybridisation analysis. In the OFP technique partial sequence information is achieved by measuring the hybridisation of a limited number of short DNA oligoprobes of with targets. By extending the number of oligoprobes to

all possible n-mer (4^n), it is possible to derive the nucleotide composition of only given fragment of DNA from the analysis of hybridisation signals. In fact, this method requires absolute discrimination of perfect hybrids from those containing mismatches.

For oligoprobes with 11-20 nucleotides in length, the frequency of complementary target in DNA fragments of interest remains too low to allow practical sequence fingerprinting. However, for oligoprobes 5-10 nucleotides in length, there is an exponential increase in the hybridisation frequency. Decreasing oligoprobe length permits the use of reasonably smaller number of oligoprobes to provide sufficient sequence information for large set of unknown DNAs.

Calculations of average hybridisation frequencies of randomly selected oligoprobes to random dsDNA sequences which are independent from each other with a length of ca. 1 kb, provided the following numers: octamers -3%, heptamers -11%, hexamers -39%, and pentamers -86% (Herwig et al. 2000). Due to the higher hybridisation frequencies of 7-mers and 6-mers, the use of these oligoprobes, compared to 8-mers, was expected to deliver higher partitioning and hence better clustering of clones of interest (Herwig et al. 2000). This increase if partitioning through reduction of the oligoprobe length was one of the motivatives of this project.

8.1.1.2 Shorter oligoprobes are cost-effective

The principle of OFP relies on the establishment of DNA-specific patterns of hybridisation with a set of oligoprobes of defined length so that similar dsDNA clones can be grouped into the sequence clusters. Obviously reduction of the oligoprobe length results in a significantly reduced number of hybridisations necessary for production of the unique fingerprint. For example, only 70 hexamer are sufficient for a meaningful clustering, whereas for heptamers 100 oligoprobes are recommended (Bauer et al. 2004). Application of the octamer requires using 220 oligoprobes (Herwig et al. 2002). Reduction of the number of hybridisations results, in turn, in a drastic reduction of costs in terms of labour, reagents and time consumption. And because we have used an extremely short LNA-modified oligoprobes, the generation of a universal set of genotyping oligoprobes is possible.

8.1.2 Usage of LNA

8.1.2.1 Incorporation of LNA modification into oligoprobes

Reduction of DNA oligoprobe length is not always possible because of limitations related to melting temperature (T_m). The T_m of short oligoprobes is often significantly reduced, which prevents hybridisation. Locked nucleic acid (LNA) is a new class of DNA analogue whose incorporation into oligonucleotides results in a significant increase in the thermal stability of duplexes with complementary DNA (Orum et al. 1999; Jacobsen *et al.* 2002).

LNA stacking in the heteroduplex results in enhanced duplex stability, even for very short oligoprobes (Simeonov and Nikiforov, 2002). The melting temperatures (Tms) of some selected non-modified and LNA-modified oligoprobes were predicted in this study by an online program as described in 'Methods' and listed in Table 6. Based on these data, it was clear that Tms equivalent to 8-mer or 7-mer DNA oligoprobes can be achieved with shorter LNA-modified oligoprobes (7-mer or 6-mer) by incorporating only 1-2 LNA monomers. Hence a window for more sensitive single-base mismatch discrimination could be opened for genotyping assays.

Table 6. Tms of LNA-modified oligoprobes equivalent to non-modified ones with one more nucleotide

Name	Sequence	Tm (^oC)	Name	Sequence	Tm (^oC)	ΔTm (^oC)	LNA Nr.
OP-01L-6nt	cAgaAg	-2	OP-01	tcagaag	-5	3	2
OP-17L-6nt	tgCTgg	9	OP-17	tgctggt	3	6	2
OP-19L-6nt	tGccAa	0	OP-19	ttgccaa	0	0	2
OP-13L	cctccTg	9	OP-13-8nt	tcctcctg	10	-1	1

Modifications with LNA bases are indicated with capital letter in the oligoprobe sequences. The letter 'L' denotes LNA-modified oligoprobe.

The Tms of seven pairs of DNA and LNA-modified heptamers, as well as one pair of DNA and LNA-modofied hexamers were predicted according to 'Methods' and listed in Table 7. It was observed that increases in melting temperature per LNA modification (ΔT_m values) is in the range of 2~8℃, in concert with the conclusion of 1~8℃ (Obika *et al.* 1998; Wengel, 1998). Especially for very short (6-mer) oligoprobes, the impact of introduction of LNA monomers was most pronounced, as

the most significant Tm gain was 8℃ for OP-06L-6nt versus OP-06-6nt. Based on the observation that OP-02A (cTgaagc, one LNA modification on T) and OP-02B (CtgaaGc, two LNA modifications on G/C) have the same Tm value (7℃), we hypothesize that LNA modification on A/T can result in more significant Tm increase than that on G/C since G/C itself has strong binding affinity.

Table 7. Tm comparison of LNA-modified and corresponding non-modified heptamers and hexamers

Name	Sequence	Tm (^{o}C)	Name	Sequence	Tm (^{o}C)	ΔTm (^{o}C)	LNA Nr.	ΔTm/LNA (^{o}C)
OP-01L	tcAgaAg	6	OP-01	tcagaag	-5	11	2	5.5
OP-02L	CtgaaGc	7	OP-02	ctgaagc	3	4	2	2
OP-03L	aTgAgGa	13	OP-03	atgagga	-7	20	3	6.7
OP-13L	cctccTg	9	OP-13	cctcctg	4	5	1	5
OP-15L	cTcctCc	12	OP-15	ctcctcc	3	9	2	4.5
OP-17L	tGCtgGt	16	OP-17	tgctggt	3	13	3	4.3
OP-19L	ttGccAa	9	OP-19	ttgccaa	0	9	2	4.5
OP-06L-6nt	Tcctcc	-5	OP-06-6nt	tcctcc	-13	8	1	8

Modifications with LNA bases are indicated with capital letter in the oligo sequences. The letter ‘L’ denotes LNA-modified oligoprobe.

The discrepancy between the number of LNA modifications in oligoprobes and the hybridisation specificity was observed. As demonstrated for hybridisation on nanoporous membrane slide in Fig. 26, the presence of more than two LNA modifications resulted in decreased specificity. On the other hand, the oligoprobes with three LNA modifications, as for example oligoprobes OP-03L, -05L and –17L in Figure 35, were sufficiently specific. From this observation we conclude that either two or three LNA modifications in heptamer oligoprobes will lead to improvement in hybridisation specificity.

8.1.2.2 Better mismatch discrimination with LNA-modified oligoprobes

Shorter oligoprobes should give better mismatch discrimination, since the relative decrease in hybrid stability with a single end mismatch is greater than for longer oligoprobes. The sensitivity of LNAs in mismatch discrimination makes them particularly attractive substitutes for conventional DNA oligoprobes in hybridisation-based analysis. The data in Fig. 18 show that hybridisation signal of 6-mer LNA-

modified oligoprobe (OP-06L-6nt) with its matching ssDNA target (DNA-2) is much more discriminative from that with the mismatching ssDNA target (DNA-1) which is on the same signal level of unrelated ssDNA target (DNA-4) and background (without any DNA target). Whereas for 7-mer and 8-mer DNA oligoprobes (OP-27 and OP-13-8nt) the signals from mismatch binding are up to more than 35% of those from matching duplex, resulting in difficult discrimination. And single-nucleotide mismatch discrimination is better for LNA than for DNA, which we could see from the hybridisation performances of oligoprobes of OP-06L-6nt and OP-06-6nt.

8.1.3 Oligoprobe pre-selection

The oligoprobes should be informative for the clone sequences in the sense that all different genes can be distinguished by their fingerprints. This implies that oligoprobes should hit the clone sequences with a considerable frequency. The simple pre-selection of oligoprobes according to high frequencies might lead to the accumulation of oligoprobes that are highly similar to each other so that the gain in information about the clone sequences does not significantly increase. Among our 26 randomly selected oligoprobes, 5 pairs of oligoprobes (OP-02L and OP-04L, OP-05L and OP-17L, OP-06L and OP-15L, OP-14L and OP-23L, OP-24L and OP-25L, Table 1) had high degree of sequence similarity, most with one-base difference at the end (except for the pair OP-14L and OP-23L). This could be a reason for discrepancies between our experimental clustering and the in-silica clustering (Fig. 39 and Fig. 40), i.e. not all of clone members for each cluster fall into the distinct group. For example sequence clusters 1 and 2 were spit into subgroups. However, as demonstrated in Fig. 32, the oligoprobe pairs with high sequence similarity having only single-base difference at the end (OP-24L and OP-25L) could be applied to confirm the hybridisation specificity.

8.2 Liquid-based hybridisation assay in miniaturised array platforms

8.2.1 Non-radioactive detection

In order to apply the OFP approach efficiently, many clones have to be analysed in parallel. The use of radioactively labeled oligoprobes as in conventional OFP application restricts the array density to about 125 clones/cm^2. Higher spotting densities would provide poorer data quality because of spot-to-spot overshining

effects. In general the use of radioactive materials requires expensive laboratory infrastructure and it is hazardous. Another limitation of this methodology is the short life of the radioactively labeled oligoprobes.

In contract to advances in fluorescent labeling of nucleic acids and fluorescent detection have made the use of arrays simpler, safer and more accurate. One of the advantages when using fluoresce is that it overcomes the need to use radioactive isotopes. Avoiding radioactivity and introducing nonradioactive detection has several inherent benifits. First, neither health risks nor waste disposal problems are encountered. Second, the longer stability of oligoprobes implicates more convenient planning of experiments. Third, no special laboratory facilities are required.

Fluorescently labeled oligoprobes in combination with laser scanner or CCD-camera detection can dramatically reduce the detection time and exclude signal interference, as long as the laser to spot diameter is smaller than the spot-to-spot distance. Replacement of radioactivity by the fluorescence labeling method proposed here facilites parallel analysis of a large number of clones.

Fluorescence permits the possibility to detect two or more different signals in one single experiment. This makes multiplexing measurements possible for comparative analysis of multiple samples, and has also increased the accuracy and throughput of analysis. Fluorescence-based assays are ideal for high-throughput, miniaturised screening systems due to high sensitivity. Hence, all of the experiments in this study were conducted using fluorescently-labeled (e.g. Cy5) oligoprobes.

8.2.2 Cy5 as the acceptor dye

For each fluorescence platform, an appropriate dye or combination of dyes had to be selected. Howell et al. (2002) used Sybr Green I as a FRET donor, together with ROX (6-rhodamine) or Bodipy TMR as acceptor dyes. Here, the range of possible acceptor dyes was extended by Cy5. Full characteristics of the dyes can be found at the Molecular Probes website (http://www.probes.com).

The advantage of the organic dyes with emitting wavelength in the far-red (for example >650 nm), particularly the cyanine dyes (Cy5, Cy5.5 and Cy7), is the considerably reduced background fluorescence (Selvin, 2000), which makes cyanine dyes of choice in iFRET assays (Ha et al. 1996).

Despite apparent insufficient overlap of Cy5 spectra with Sybr Green I emission (Cy5 absorption maximum is at 646 nm, and Sybr Green I emission maximum is at

518 nm), our results show that Cy5 could be used as an acceptor in an iFRET pair with Sybr Green I. This is probably due to the fact that the Cy5 dye, when attached to DNA, acquires additional shorter wavelength absorption peaks compared with the spectrum of the free dye (Molecular Probes Handbook, 2005). Having tested various combinations, shown in Fig. 15 and Fig. 16, we revealed that, for our iFRET detection set-up, Cy5 is the best acceptor dye for Sybr Green I. It provided a high iFRET signal, as well as a good spectral separation of the signal from the excitation dye peak. It was therefore possible to monitor both signals with appropriate filters.

8.2.3 SYBR Green I as the donor dye

SYBR Green I can be excited with blue light with a excitation wavelength at 480 nm and maximum emission spectrum at 520 nm. It is the most frequently used dsDNA-specific dye. The increase of fluorescence upon intercalating into dsDNA is more than 1000-fold higher than that of the free dye. This dye is therefore well suited for monitoring the dsDNA accumulation (Wittwe et al. 1997).

Even though the dsDNA-specific dyes YOYO-1 and TOTO-1 are also asymmetric cyanine dyes and structurally related to SYBR Green I, a shift to longer-wavelength emission and a concomitant drop in quantum yield of the TOTO-1 and YOYO-1 dyes are observed upon binding of these dyes to single-stranded nucleic acids (Molecular Probes Handbook, 2005). Furthermore, the fluorescence intensities of these two dyes tend to decrease gradually because of bleaching by continuous laser irradiation (Yoshihiro et al. 2004). Hence, the dsDNA-specific dyes YOYO-1 and TOTO-1 were excluded from the consideration as the dsDNA-intercalating dye for our iFRET system.

Because of its high specificity for dsDNA, its bright fluorescence and low cost, SYBR Green I was chosen as the intercalating dye for this study. Its stabilizing effect on DNA duplexes, resulting in an effective increase of its melting temperature (Prince et al. 2001), should also be advantageous for use of very short oligoprobes such as 6-mer oligoprobes used in this iFRET study.

8.2.4 Solution-based miniaturised assays

Homogeneous ‘mix and measure’ assays in miniaturised formats are well suited for high-throughput screening (HTS) as they do not require separation and wash steps, which are time-consuming and difficult for automation. Together with the advent of miniaturised technology, the fluorescence-based approaches for detecting the molecule–molecule interactions become increasingly popular in homogeneous assays. Moreover the use of high-density micro-well plates together with plate-handling robotics enabled simultaneous analysis of thousands of variables in a single experiment.

Technology development led to conversion from a limited number of reactions in a greater volume to an extremely large numbers of assays performed in nanolitre and picolitre volumes. In this study we applied nano-dispensing techniques (TopSpot in Fig. 6 and Fig. 7, and sciFLEXARRAYER in Fig. 9). Thus miniaturisation of the hybridisation process was achieved (Fig. 11). However, downsizing of reaction volume from 100 µl solution in cuvette (Fig. 17A) to 2 µl droplet on glass slide (Fig. 19), and further to 30 nl liquid in a re-usable nano-structured high-density plates (Fig. 23) could be established. The solution-based miniaturised approach presented here could potentially contribute to a significant increase in the throughput of large-scale genomic applications, such as OFP and genotyping. This is an enormous advantage for oligoprobe/sample analysis that involves a high-throughput, such as OFP (Meier-Ewert et al. 1998; Herwig et al. 2000, 2002), in which a minimum of 100,000 clones have to be probed individually with at least 100 oligoprobes, which means 10^7 independent assays.

But miniaturisation of biological assays is more complex than just transferring reactions to smaller volumes (Bilitewski et al. 2003). A few fundamental challenges need to be addressed (Stahl, 1999). Firstly, as the dimensions decrease, the surface to volume ratio becomes higher through miniaturisation, making surface treatment important in preventing non-specific binding of components to the system walls. Secondly, with miniaturisation, the small volume of the assay raises the concern of the effect of evaporation on assay quality. Thirdly, fluid containing ions can poison the devices and render them useless. Fourthly, by reducing the volume of the assay, the influence of small molecular differences increases. Reactions within small volumes will be limited by the diffusion of the components (diffusion limit). This problem can be overcome by active mixing of the liquid.

8.3 Nanoporous membrane-based hybridisation assays

8.3.1 Nanoporous membrane as the support

Fluorescence detection provides low-level signal intensities in comparison to radioactive detection. The autofluorescence of classical nylon membranes results in high background signals. In order to overcome these obstacles, at least partially, we applied the Pall nanoporous nylon membrane slide in microarray-based hybridisation assays.

Arrays on nanoporous membranes have been in use for many years (Southern et al. 1992, 1994). Membrane-based cDNA array is an attractive economical alternative for high throughput gene expression assays. To this end, membranes (Chen et al. 1998) were covalently bound to glass surfaces to improve hybridisation signals as compared to the limited oligoprobe binding capacity of two-dimensional (2-D) surfaces, for example glass slides.

Three-dimensional (3-D) surfaces have a greater number of potential immobilization sites than modified glass. The amount of oligoprobe is increased due to the larger surface area provided by the 3-D structure. This is probably the main reason for the elevated hybridised density (Chen et al. 1998) and better data consistency (Livshits and Mirzabekov, 1996; Afanassiev et al. 2000). This property can lead to semi-quantitative retention of arrayed DNA (manifested in Fig. 27), resulting in a higher concentration of DNA per spot than on glass. The 3-D structures are also providing a solution-like environment for hybridisation (Hong et al. 2005). Furthermore, the nucleic acids can be applied in a relatively large volume since they soak into the pores of the membrane. The relative fragility of this material is not a problem given the small size of the arrays. Membrane-based technology has also one advantage of being well established within scientific community.

8.3.2 Re-utilisation of microarrays

A major cost in microarray experiments is the microarray itself. Thus, the ability to reuse the microarray is of significant advantage. The number of stripping and reprobing cycles that a membrane can withstand is an important economic factor when selecting a membrane. Re-usability of microarrays helps to eliminate the variance between arrays and can rescue an experiment that failed due to high non-specific background.

Hence, repetitive experiments with the same set of membranes can be performed, and the averaging of OFP data leads to a higher statistical significance. This requires an effective membrane-stripping protocol prior to re-hybridisation with a different oligoprobe. The commonly used boiling of the membrane in an SDS solution limits the re-use to 5 times because of physical damage to the membrane and a reduction of spotted cDNA on the membrane surface (BD Biosciences Clontech Atlas‘ cDNA Expression Arrays User Manual, PT3231–1). It has been shown, that by stripping with standard NaOH solutions, the same nylon membrane-based microarrays could be re-used for more than 10 times without a measurable reduction in their performance (Bancroft et al. 1997; Hornberg et al. 2002). After this point the surface of the filter begins to suffer from mechanical damage. To minimise the deterioration during the stripping process, a gentler stripping protocol is needed.

In this study, Pall nanoporous membrane could be stripped and re-probed for at least 19 times without losing significant signal intensity as described in shown in Fig. 28. Our protocol allowed for more comparative studies of multiple data sets derived from sequential hybridisations of the same set of membranes.

8.3.3 Hybridisation conditions

In practice a certain amounts of oligoprobes will not perform with a suggested rate of 15-20%, due to unfavorable hybridisation conditions (Herwig et al. 2000). There could be two possible reasons for low hybrid formation: 1) prevention of hybrid formation and 2) hybrid instability. The inability to form a hybrid can be inherent to the oligoprobe (self-complementarity) or the unavailable target (presence of secondary structures).

DNA sequences form stable duplexes only in the presence of salt required for counteracting the inter-strand repulsion. Such conditions, however, stabilise secondary and tertiary structures within a target molecule. Sequences might thereby not be accessible and be prevented from hybridisation to the gridded DNA (Southern et al. 1994). Intra- or inter-molecular structure of the oligoprobes can compete with oligoprobe-target duplex formation and result in low hybridisation intensity. Extensive secondary structure of the target can also limit this efficiency.

Thermal stability curves for very short oligoprobe hybrids (6-nt and 7-nt) are at least 15 degree lower than for hybrids with 11-12 nt oligoprobes. 6-mer or 7-mer duplexes have a low stability and must be hybridised at low temperature. Performing

the hybridisation reaction at optimal conditions, e.g. at the lowest practical temperature 4℃, time and empirically optimised oligoprobe concentration allowed us to maximise the hybrid formation and efficiently detect complementary sequences in unknown DNAs.

Extended washes of the resulting hybrids at the same temperature (4℃) maximise discrimination by allowing the higher dissociation rate of mismatched hybrids. It could be potentially beneficial to perform washing at various temperatures according to the predicted stability of the duplexes. In order to simplify the experimental procedures, this possibility was excluded from our consideration. Varying the time of post hybridisation washes can also increase the specificity, but the effect on the hybrid yield is sequence-dependent. For example, for most of the 26 LNA-modified oligoprobes, slides were washed with hybridisation buffer at 4℃ for 2 min. But to obtain good specific hybridisation signals, for the oligoprobes OP-05L, OP-16L and OP-25L, washing time was extended to 5 min whereas for the OP-18L to 10 min due to the stronger affinity to target DNA as all four oligoprobes have elevated GC content (71% or 86%, see Table 1).

8.3.4 Manufacturing DNA microarrays from unpurified PCR products

DNA microarrays are currently widely used in biological and biomedical research. Purification of PCR products prior to spotting is a most tedious step. Significant amount of material is lost during purification and contamination might occur. Therefore we aimed to overcome tedious purification before deposition onto the nanoporous nylon membranes.

Our data (Fig. 33) showed that the percentage ratio of signal intensity from un-purified full matching clones is about 48-93% of the intensity from the corresponding purified molecules, which coincide with the observation of Chen et al. (1998). When un-purified PCR products were used directly as spotting solution, no increase in the clogging of spotting pins was observed. Thus formation of salt crystals in the pins during the spotting process was avoided.

In spite of 5-fold molar excess, the presence of the primer molecules in the unpurified PCR samples did not increase the non-specific signals in the spots upon hybridisation. This indicates that binding of the short (24 nt) primer molecules to the nanoporous nylon membrane is a relatively rare and random event. It has long been recognised that short DNAs (<100 bases) do not interact with membrane as

efficiently as long DNAs (>100 bases) (Van Oss et al. 1987). However, even if a small number of primer molecules are still bound, they will not significantly contribute to the signal intensities, due to their lower duplex stability and kinetic disadvantage compared to the PCR products.

This signal loss (7-52%) of hybridisation, however, was offset by the variation in the amount of DNA present at the individual spot positions across an array due to the system errors if the considerable savings in time and cost of the experiment was not taken into consideration. In addition, a larger number of arrays can be made from one batch of amplification products.

8.3.5 Spot shape and geometric evaluation

Image analysis is a major part of data evaluation for array hybridisation experiments. Few microarray images are perfect. Many microarray distort (e.g. overshining, scratch and comet-tails) that we have encountered, as shown in Fig. 34, are difficult to trace, and do not manifest until the hybridisation is finished.

One common spot shape problem is formation of comet tails on the array (Fig. 34F). This phenomenon is likely caused by too slow immersion of the slides into the blocking solution. Another problem is that spot homogeneity depends on the variation of the DNA concentration across a single spot, such as a higher DNA concentration at the edges (so-called ‘doughnut’ effect). Once deposited, DNA is not distributed uniformly over the spot area while much more material assembles near the outer edge of the spot with a relatively empty area in the centre. This phenomenon could usually be alleviated by longer re-hydration. However, comet-tails or donut holes did not make an array unusable and in fact data quality can be restored. This was achieved by reducing the fluorescent area being analyzed or manually excluding imperfect spots from the analysis.

DNA spots of high homogeneity are beneficial, since they simplify image analysis and enhance the accuracy of signal detection. Hence, the array may not have the strict geometric regularity (Fig. 34A, indicated by the arrow) with the boundaries and shapes of the spots poorly defined. Furthermore, it was not possible to reduce the volume of spots beyond certain limits (e.g., 4 nl for synthetic ssDNA demonstrated in Fig. 27 or 20 nl for PCR products illustrated in Fig. 31), or to control their size and shape on a porous membrane. Thus, it was not possible to locate spots with the high precision that can be achieved on a rigid substrate. This caused a

difficulty in developing highly accurate grids to specify target locations for automated analysis of hybridisation signals. Instead, manual analysis was required for the problematic spots, which could however lead to laborious and inaccurate outcomes.

8.3.6 Background and its subtraction

There is a distinct class of array problems related to abnormally high fluorescent background. High background was observed across the entire hybridised surface (Fig. 34D) as well as in localized regions on the array (not shown). Many potential sources of the high background include the contaminations (Fig. 34E-G), elements of the hybridisation process (e.g. container, solution, temperature and time), of the array printing process (e.g. print head status, robot movement, quality of substrate and uniformity of the membrane slides), of image acquisition (e.g. type of optics, laser cross-talk, dye quenching and laser power), of spot identification and quantification. It was however difficult to determine the source of a particular type of high background. The standard approach is to simply subtract the local background directly from the spot intensity. We measured the intensity of small regions surrounding the spot mask and subtracted it as background.

Background subtraction is a crucial element for the subsequent data evaluation. Each pixel value represents the level of hybridisation at a specific location on the slide. Usually, the observed signal is thought to be a combination of the true signal from the specific hybridisation of interest, and the background signal as a result of non-specific hybridisation and contamination. In our study the background was subtracted to obtain a more accurate quantification of hybridisation. It seems that the fluoresence outside of the DNA spot is different from that within DNA spot.

8.3.7 Signal evaluation

DNA hybridisation process often produces false negative errors (i.e. oligoprobes that match the sequence but show no or extremely weak signals) and false positives (oligoprobes display hybridisation signals but without matching any position in the sequence). The false negatives reduce the sensitivity of the OFP method by an under-estimation of the gene expression levels. False positives tend to occur when there are minor, for example 1 bp, mismatches between oligoprobe and target, particularly when the mismatch occurs at the end of the oligoprobe.

In an OFP analysis of a set of ^{33}P-labeled octamers, false negative and false positive errors were found to be randomly introduced on the level of 20% for each (Herwig et al. 2002). Our OFP study has shown that much more specific hybridisation can be achieved with the 26 LNA-modified heptamers. The false negative and false positive errors for each individual LNA-modified 7-mer oligoprobe were found in the range of 0-12.12% and 0-16.67%, respectively, and with the average 3.03% and 4.25%, respectively. Compared with the DNA octamers mentioned above, the significant improvement of hybridisation performance of heptamers carrying LNA modifications might attribute to the increase of Tm (LNA intercalating) or binding frequency (shorter oligoprobes).

On the other hand, hybridisation signals vary across the same hybridisation from one single oligoprobe (OP-24L in Fig. 32A) or between hybridisations from different oligoprobes (OP-01L and OP-05L as shown in Fig. 29). Apparently the hybridisation characteristics of an oligoprobe with a fixed length (e.g. 7-mer) depends on the nucleotide content of the oligoprobe, the distribution of similar sequences in the target, the dye attached to the oligoprobe and the hybridisation conditions.

According to the experimental steps listed in Methods we composed an uncomplete list of the major possible sources of false positive/negative signals and signal fluctuations. The list addresses fluctuations in oligoprobe, target and array preparation, hybridisation process, background effects and image processing:

A. DNA quantification. Clones are subject to PCR amplification, which is difficult for accurate quantification using Bio-Photometer. Hence even though the concentration of all the 66 genomic DNA was normalised to the same 0.1 µM, the concentration variation will lead to the signal fluctuations for the positive dsDNA across the same single slide.
B. Dye bias. This can be caused by the efficiency of Cy-5 incorporation during oligoprobe synthesis. Guanine in the vicinity of the fluorophore can reduce the fluorescence. An oligoprobe with many cytosine and thymine bases has only 60% of the Cy-5 absorbance of an oligoprobe containing only guanines and adenines (Olfert Landt, personal communications). Illumination can lower the fluorecence intensity of Cy5-labeled oligoprobes and the stability is lower for diluted solutions. Therefore more attention should be paid to oligoprobe design and synthesis as well as the storage.

C. Random fluctuations in target volume. The amount of transported target fluctuates stochastically even for the same nozzle.
D. Target fixation. The fraction of target cDNA that is chemically linked to the membrane is unknown. A minimum number of bases are needed to interact with the immobilisation substrate. Bases that are involved in the interaction with the surface are unavailable for Watson–Crick base pairing with solution-phase DNA (Franssen-van et al, 2002). Longer DNA strands require a smaller proportion of bases for attachment and therefore have a greater number of hybridisation sites and ultimately generate higher signal intensities (Stillman and Tonkinson, 2001).
E. Hybridisation parameters. Hybridisation efficiency was influenced by a number of experimental parameters, such as temperature, time and buffer conditions.
F. Non-specific background and hybridisation. Non-specific radiation or signals from neighbouring spots cannot be completely excluded.
G. Image analysis. Non-linear transmission characteristics also distort the hybridisation signals. Saturation effects and variations in spot shape can cause some problems.

8.3.8 Clone clustering

After the data normalisation, fingerprints of all clones were compared to each other, and a similarity score could be obtained for all possible pairs. The aim of the clustering procedure is to classify the clone fingerprints according to a well-defined pair-wise similarity to group similar fingerprints together and then to separate dissimilar ones. Clustering results help normalise cDNA or genomic libraries and thus significantly reduce sequencing effort in gene identification. The calculated classification reflects the number of different genes expressed in the tissue (number of clusters) and their relative abundance (size of clusters). When processing tissues from different developmental stages, clustering can detect differences in gene expression and thus identify development-specific genes.

However, there are many sources of systematic variation in microarray experiments. Since the actual signal intensity depends on a number of parameters, which are often difficult to control as discussed above (amount of DNA in each spot, exact hybridisation conditions and sequences surrounding the match site), the identification of clusters of similar or identical clones by OFP can still in many cases

be incorrect. Clones from the same transcript can be clustered into a few separate clusters. Furthermore individual clones belonging to a cluster can be left as singletons or assigned to wrong clusters.

Due to the noise in hybridisation signatures, a single group of identical clones may be split across several clusters, causing under-clustering. For example, in our model study (Fig. 40), clusters 4, 5 and 7 were reconstructed with only one clone (D18, J19 and I17, respectively) assigned as singletons, and cluster 1 and 2 were spread into subgroups. For cluster 1, clones F12, F18 and H02 were divided separately as singletons from the main cluster. For cluster 2, the 15 clones were separated as four sub-clusters. Even though we used a k-means-based algorithm and sequential rounds of clustering (Herwig et al. 1999) in order to partially alleviate the under-clustering, it persisted in our investigation.

The splitting of clones representing transcript obviously leads to an over-estimate of the complexity of the expression profiling, i.e. under-clustering. In some cases, this is likely to be caused by remaining imperfections in the image analysis software, such as cross-talk from a neighbouring clone with strong hybridisation. Additional source of under-clustering may be presence of alternative splice forms that represent different mRNA species and alter clone fingerprints. These could be correctly split into different OFP clusters, yet be from the same gene.

Along with the under-clustering phenomenon, over-clustering also takes place, which means that cDNAs from different genes are placed in the same cluster. For instance, in Fig. 40, clone D18 from cluster 4 was wrongly aggregated to cluster 2. I17 from cluster 7 and J19 from cluster 5 were separated as singletons.

Small clusters are harder to identify because of the high variance introduced by experimental error. The identification of small clusters requires very stringent set-up of algorithmic parameters. This leads to overestimation of the total number of genes due to cluster splitting and to the false assignment of clones to singletons that should be clustered and thus to a lower normalisation rate. A straightforward way to overcome this problem is the use of more oligoprobes.

The fingerprints were obtained upon hybridisation in this study from short oligoprobes (8- and 9-mers, or even 7-mers) and should be different for different clones even if they contain the same repeat elements. This is because the repeat covers only a small fraction of the clone sequence. Our result shows that the OFP approach well reproduces the theoretical sequence clusters. Separated and highly

overlapping clusters such as sequence clusters 3, 6, and 8 were recovered accurately (Fig. 40). Even if some clones (e.g., D18, F12, F18, H02, J19 and I17) were assigned to wrong clusters, they exhibit low similarity of their individual fingerprint to the consensus fingerprint of that cluster and will not be selected as representatives of those clusters for sequencing.

8.4 Development of cost-effective hybridisation approaches

8.4.1 In-solution iFRET hybridisation technology

One aspect of our work was to minimise the cost of screening. This was addressed primarily by reducing the sample volumes. Current technology development leads to conversion from a limited number of reactions in a greater volume to extremely large numbers of assays performed in nanolitre and picolitre volumes. This is an enormous advantage for oligoprobe/sample analysis in a high-throughput such as OFP (Meier-Ewert et al. 1998; Herwig et al. 2000, 2002). In OFP a minimum of 100,000 clones have to be probed individually with at least 100 oligoprobes, which means that 10^7 independent assays have to be performed. The traditional OFP method, which is based on radioactive oligoprobe hybridisations using DNA clones that have been immobilised on nylon membrane filters, is limited by the high experimental costs and time consumption. The conceived nano-well format, combined with an one-step protocol of fluorescent hybridisation detected by a sensitive CCD camera, should reduce the experimental load by a factor of 10^3, reducing the load from 10 µl to 10 nl.

8.4.2 On-surface nanoporous membrane hybridisation technology

Without consideration of set-up cost, the cost for conventional nylon membrane hybridisation for 100,000 clones using 250 radioactively-labeled 8-mer normal DNA oligoprobes is estimated to be 3 Euros per clone, with 4 persons in 3 months for the OFP project (Uwe Radelof, personal communications). The cost of nanoporous nylon membrane hybridisation approach established in this study, using 100 Cy-5 labeled LNA-modified heptamers applied to 100,000 clones is calculated to be less than 2 Euros per clone, with 2 persons within the period of 2 months. Assuming that the price of generating LNA is expected to decrease in the future, the cost of this new approach is supposed to decrease further.

8.5. Outlook

Theoretical calculations predict that 110 heptamer hybridisation oliogoprobes should be sufficient to generate unique fingerprints, whereas in our study only 26 LNA-modified heptamers (Table 1) and 5 LNA-modified hexamers (Table 2) were used. More hybridisations will be required in order to cluster a larger collection of DNA clones. Taking into account the experimental noise and the fact that not all fingerprints are equally significant due to statistical variations, it is clear that the sensitivity of the method can be improved by using more oligoprobes. Increasing the number of oligoprobes would mean that smaller differences in sequences could be discriminated. It should be even possible to detect splice variants of a single gene using this approach. Due to the inaccuracy of hybridisation experiments, the fingerprints are identified with certain degree of error. Most frequently hybridisation error results in single-base mismatch at either end of oligoprobe. The error rate decreases rapidly with more hybridisation experiments (Milosavljevic, 1995).

8.5.1 Design of LNA-modified oligoprobes

The accuracy of OFP hybridisation needs to be increased by selection of oligoprobes with higher hybridisation specificity, or improvement of hybridisation conditions to increase specificity. It is not currently possible to accurately predict the specificity of hybridisation for short oligoprobes. We should select sets of more specific oligoprobes by empirically assessing the quality of hybridisations through the use of control clones with known sequence.

Ideally, at the hybridisation temperature at 4℃ used in this study, it should be possible to discriminate between perfectly matched oligoprobe-target duplexes and mismatched duplexes as results of the differences in their thermodynamical stabilities. In this setting, the increased discrimination power and specificity of LNA-modified oligoprobes may provide more informative results. LNA bases could be selectively incorporated to fine-tune thermal stability of the oligoprobes. Oligoprobes with low G•C content could contain more LNA residues than oligoprobes with high G•C content. The number of LNA residues could be adjusted, so that various oligoprobe-target duplexes could achieve similar stability.

It has been demonstrated (Koshkin et al. 1998) that LNA–DNA base pairs are more thermodynamically stable than DNA–DNA base pairs. It is therefore likely that problems arising from oligoprobe dimers will be more significant when LNA residues

are present. However, 7-mer or 6-mer oligoprobes used in this study were not prone to form self-dimers.

LNA effects vary with the identity of the mismatch, length and composition of the oligoprobe, sequence context, LNA modification number and position. Generally, LNA substitution gives a smaller ΔT_M in longer oligoprobes. The largest impact upon introduction of LNA nucleotides is observed for short oligoprobes with one or more centrally positioned LNA monomers (Petersen and Wengel et al, 2003; McTigue et al. 2004). Furthermore, separated single or multiple LNA modifications appear to have a larger impact than contiguous stretches of LNA nucleotides.

LNA pyrimidines contribute to higher stability of duplexes than LNA purines (in order: LNA-C > T > G ≫A), but there is substantial context dependence for each LNA base. Both the 5' and 3' neighbours must be considered in predicting the effect of one LNA incorporation, with purine neighbours providing more stability (McTigue et al. 2004).

8.5.2 Exploitation of Alexa Fluor 647 and EvaGreen dyes

Even though cyanine dyes such as Cy3 and Cy5 have been widely used in many applications due to their long excitation wavelength and low background fluorescence, the fluorescence bleaching (Gruber et al. 2000) still limits their usefulness. As a consequence, we observed some common signal variations as discussed in section 8.3.7. Hence we need to exploit some alternative dyes to improve the hybridisation performance in the future. Some reasons are listed below.

Nucleic acids labeled with Alexa Fluor dyes show brighter fluorescence than those labeled with similar dyes. The Alexa Fluor dyes have several advantages to make them superior to other fluorescent dyes: 1). The high water solubility of nucleic acids labeled with the Alexa Fluor dyes, even in high-salt conditions, makes them ideal for hybridisation experiments. 2). The fluorescence of the Alexa Fluor labeled nucleic acid is not pH sensitive within the ranges used for hybridisation. 3). The resistance to photobleaching of Alexa Fluor dyes would allow more times for repeated viewing of labeled samples and image capture hence allowing them ideal for applications such as FRET, FISH and microarrays.

Alexa Fluor 647 and Cy5 dyes are structurally distinct dyes that have similar spectral characteristics: both dyes display absorption maxima at approximately 650 nm and emission maxima at approximately 670 nm. Cy5-labeled DNA exhibited

reduced fluorescence and absorption anomalies, while Alexa Fluor 647-labeled DNA exhibited little or no such anomalies (Molecular Probes, handbooks). The functional benefit of these properties of the Alexa Fluor dye is seen in better signal correlation in microarray-based hybridisation assays (Cox et al. 2004) and should be investigated in iFRET setting.

EvaGreen is a novel dsDNA intercalating dye that is more stable and sensitive than SYBR Green I (according to the manufacturer's product and safety data sheet for EvaGreen, available on http://www.biotium.com/product/product_info/allcolor.pdf). Upon binding to DNA, the fluorescence of EvaGreen is several-fold higher than that of SYBR Green I. Unlike SYBR Green I, which has been reported to be unstable, EvaGreen is highly robust, both thermally and hydrolytically under alkaline or acidic condition. In addition, the absorption and emission spectra of EvaGreen are similar to those of SYBR Green I, which means that the same optical setting for SYBR Green I can also be used for EvaGreen. Our initial data (not shown) have demonstrated that when SYBR Green I was replaced with EvaGreen using the same concentration for iFRET hybridisation in solution, more significant and discriminative signals (between fullmatch and mismatch, between signal and background) were observed.

Taken together, we concluded that, the usage of Alexa Fluor 647 labeled oligoprobes should greatly improve nanoporous membrane hybridisation, and the iFRET system along with the usage of intercalating dye EvaGreen.

8.5.3 Trends in miniaturisation: microplates versus microfluidic chips

The recent development of microfabricated fluidic devices has the potential for triggering revolution in HTS technology, and it is currently being actively pursued in both academic and industrial laboratories. Microfluidics is an emerging technology which allows the movement of minimum volumes in microscopic channels and chambers, and enables the processing of minute volumes of liquids for chemical, biochemical, or enzymatic analyses. The merging of microfluidics and microarray technologies constitutes an elegant solution that will automate and speed up microarray hybridisation. These devices make use of electro-kinetic pumping and hydrodynamic pressure to control nanoliter-scale fluid flow. They offer the potential for achieving orders-of-magnitude reduction in reagent consumption, parallel sample processing for higher throughput, implementation of unique assay read-outs, and improved data quality.

Microfluidic device (MFD) technology – typically used to transfer small volumes of both liquids and gases – possesses great analytical potentials for the future and provide a route for ‘total’ lab-on-a-chip (TLOC) systems. It is a dynamic branch of analytical science with huge possibilities for bio-analysis (Khandurina and Guttman, 2002). MFDs will not simply be used as alternatives to microarrays but will be fused with them to produce fully integrated modular microfluidic microarrays. Possibly the most assured application of microarray and MFD hybridisation will be the creation of specific and adaptable ‘micro total analytical systems’ (µTAS), for example, this newly emerging MFD technology can be applied to our iFRET platform.

A number of improvements, in terms of fine-tuning the oligoprobe design, dye exploitation, the application of fragmented ssDNA as hybridisation target for iFRET detection system, more advanced clustering algorithm, speeding up and down-sizing the hybridisation system, still need to be made in order to increase the sensitivity of the method described in this study. The proposed approaches, with application of 7- and 6-mer LNA-modified oligoprobes in hybridisation assay might be used as a robust platform for in-solution or on-surface miniaturised, genome-wide clone characterisation suitable for automation in high-throughput applications.