Cranfield University

Clair Gallagher

Title: Multiplex PCR Suspension Array Method Development for High-Throughput NIDDM Disease Association

Cranfield Health

PhD Thesis

April 2009

Cranfield University

Cranfield Health

Doctor of Philosophy

April 2009

Clair Gallagher

Development of an automated identification system for nanocrystal encoded microspheres in flow cytometry.

Supervisors:
Prof. David Cullen and Lee Larcombe.

April 2009

This thesis is submitted in partial fulfilment of the requirements of the Degree of PhD.

Aim: This work sets out to use haplotype-based tagSNP selection and a systematic *in silico* analysis for design of multiplex-compatible PCR primer and SAT probe sets to capture maximum variation with minimum tests across candidate genes IGF1, IGFBP1 and IGFBP3. Additionally, the work aims to develop a number of robust, high-efficiency, high-specificity multiplex PCR constructs for amplification of these targets and to demonstrate the applicability of these target types to suspension array genotyping for non-insulin-dependant diabetes mellitus association facilitation.

Methods: Haplotypes for predominantly European Caucasian populations were constructed and tagSNP selection performed using Haploview to capture maximum variation across candidate genes IGF1, IGFBP1 and IGFBP3. Extensive *in silico* analysis was performed for design, evaluation and selection of robust high-specificity primer and probe pairs, suitable for downstream multiplex PCR and SAT analysis. Singleplex end-point and real-time PCR was performed for primer pair profile determination which informed multiplex PCR set construction and optimisation. The applicability of this complex target type to suspension array-based genotyping was investigated using a model probe pair using both quantum dot-encoded and fluorophore-encoded microspheres.

Results: Haploview was used for haplotype construction and linkage disequilibrium-based tagSNP selection across candidate genes, reducing the number of SNP targets from 292 to 32 with minimal information loss. Extensive evaluation of potential tagSNPs was performed and 29 SNPs, representing 29 bins across target genes were designed for multiplex analysis. Singleplex end-point and real-time PCR was performed for primer pair profile determination which allowed four multiplex PCR sets to be constructed and optimised for simultaneous amplification of 14, six, five and two targets. The applicability of this complex target type (14-plex) to suspension array-based genotyping was demonstrated using a model probe pair.

Conclusion: *In silico* analysis techniques have been applied for successful development of four robust multiplex PCR sets (14-plex, 6-plex, 5-plex and 2-plex) which display high-efficiency and target-specific amplification of tagSNPs, capturing maximum assay-compatible variation across candidate genes IGF1, IGFBP1 and IGFBP3 for European Caucasian populations. The applicability of these multiplex PCR constructs to suspension array-based genotyping has been demonstrated, thus paving the way for development of large multiplex suspension array-based genotyping assays using probes designed during the course of this work. This work offers the potential for comprehensive association analyses to become more accessible to the wider-scientific community by facilitating reduced genotyping burdens which allow increased accessibility for powerful association.

I would like to thank my supervisors Dr. Lee Larcombe and Prof. David C. Cullen for their invaluable advice, assistance and encourement during the course of this thesis.

Extra special thanks to Colin Clarke and gratitude also to the members of Cranfield Health especially Jo Holmes, Kat Low, Simon Pettifar, David Titmus and Rita and Tom for their assistance along the way and for keeping me reasonably sane. Thanks also to all former members of the dream factory and my great housemates Alex, Magnus and Claudia for providing lots of laughs this year.

This thesis is dedicated to my family; dad Charles, mum Susan, sister Lorna and brother Brian. Thanking you for a lifetime of support and encouragement.

ABCC8	apolipoprotein A2 ATP-binding cassette, sub-family C
ATP	adenine triphosphate
bp	base pairs
CAPN10	calpain-10 gene
CCD	charged couple detector
cDNA	complementary deoxyribonucleic acid
Cl	chlorine
COOH	carboxylic acid
DMSO	dimethyl sulfoxide
dNTP	Deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
DNA	deoxyribonucleic acid
EM	expectation maximisation
ESE	exon splice enhancers
ESS	exon splice silencers
FC	flow cytometry
FITC	fluoroscein isothiocyanate
FRET	fluorescence resonance energy tranfer
FSC	forward scatter
FWHM	full width half maximum
GH	growth hormone
HIV	human immunodeficiency virus
IBD	identity by descent
ID	identification
IDDM	Insulin dependant diabetes mellitus
IGF	insulin-like growth factor
IGF1	Insulin-like growth factor 1
IGF1	insulin-like growth factor-1
IGFBP1	IGF binding protein 1
IGFBP3	Insulin-like growth factor binding protein 3
IQR	interquartile range
ISS	intronic splice silencers
LOD	logarithm of odds
LTR	long terminal repeat
MCMC	Markov-Chain Monte-Carlo
MEF	molecules of equivalent fluorescence
MFI	mean fluorescence intensity
NIDDM	non-insulin-dependant-diabetes mellitus
NPL	nonparametric linkage
OLA	oligonucleotide assay
PBS	Phosphate buffer solution
PCR	polymerase chain reaction
PMT	photomultiplier tube
QD	quantum dot
QDC	quantum dot corporation
QDEM	quantum dot encoded microsphere
QTL	quantitative trait loci
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RT-PCR	real time PCR
SAT	suspension array technology
SEERS	surface enhanced Raman spectra
SNP	single nucleotide polymorphism
SSC	side scatter
T_A	annealing temperature
T_M	melting temperature
TFBS	transcription factor binding sites
TFO	triplex-forming oligonucleotides

List of figure

Chapter 1: Thesis introduction and overview

1.1 Introduction

Complex polygenic diseases such as coronary artery disease and non-insulin-dependant-diabetes mellitus (NIDDM) contribute the largest burden to ill-health care costs in developed countries [1]. Increasing efforts are being made to improve prevention, diagnosis and treatment of these costly disorders by unravelling the complex genetics which underlie disease predisposition, progression and individual response to therapeutics [2]. While monogenic diseases displaying severe phenotypes may be mapped quite effectively using smaller pedigree structures and manageable polymorphism maps, the complex nature of polygenic diseases with their multiple contributory quantitative trait loci (QTL), small effects and complex interactions including genetic heterogeneity, epistasis, low penetrance genes, pleiotropy and variable expressivity, require large population-based sample repositories and high-density maps if comprehensive analysis of target genes is to be performed effectively [3,4]. Performing association of this nature incurs incredibly large genotyping burdens and huge associated costs that limits accessibility of this technique [1] .

Many well-considered polygenic disease investigations falter due to budgetary constraints which limit sample size and breadth of polymorphism investigation rendering definitive determinations regarding candidate gene/disease associations indeterminable. If polygenic disease research is to be performed effectively care must be taken that the scope and depth of experimental design is robust enough to fully address the hypothesis

posed [1]. A number of careful experimental design innovations may be implemented to reduce genotyping burdens and render association of this nature more amenable.

Many associations to date have enabled reduced genotyping burdens by sample size reduction, however even using a best-case scenario example of a common SNP acting in a dominant fashion; in excess of 800 samples would be required to detect a strong polygenic effect with 80% power [1]. Thus for polygenic disease association, sample size reduction does not represent a viable option for genotyping burden reduction. Two techniques of considerable interest aim to effect genotyping burden reduction by use of knowledge-driven SNP prioritisation and multiplex experimental structures.

SNPs are the most frequent variation type in the human genome, occurring in every 100-300 bases; however only 1% of SNPs may be expected to confer more than modest disease associated risk, as such careful knowledge-driven selection is essential to reduce the testing burden [4,5]. SNP prioritisation may be informed using linkage disequilibrium and/or functional prioritisation-based selection which allows SNP reduction with minimal loss of power in terms of either variation coverage or putative functional alleles [5-7]. Linkage disequilibrium-based approaches are especially effective for burden reduction where homogenous populations are available and linkage disequilibrium is strong while functional prediction-based prioritisation can be effective where strong evidence (both theoretical and experimental) exists for contributory target alleles in disease predisposition [5,8].

Concomitant with SNP reduction, high-dimensionality multiplex analysis structures may be used for increased data yields. Careful design and optimisation of multiplex parameters can facilitate amplification of multiple targets in a single reaction, these concentrated amplicons may then be used for multiplex suspension array (SA) genotyping [9,10].

Suspension array analysis was first investigated in the 1970s however recent advances in multiplex capacity afforded by quantum dot-encoding mechanisms have rendered high-dimensionality combinatorial libraries compatible with standard four-colour flow cytometry and re-ignited interest in the technique [10-12]. Flow cytometers are used for a wide variety of cellular applications and as such, suspension array analysis offers the opportunity for high-throughput flexible analysis without the need for additional capital equipment expenditure. As a result multiplex PCR combined with multiplex suspension array based techniques offers an attractive alternative to expensive microarray-based work and may afford high-efficiency association to be performed in a more comprehensive fashion [13].

The motivation driving this research is to demonstrate a range of techniques that may be applied to facilitate genotyping burden reduction of NIDDM target genes IGF1, IGFBP1 and IGFBP3, and to increase accessibility of association analysis for these three candidates. It is hoped that the methods demonstrated and developed during the course of this work will contribute to association-based analyses by providing methods for robust multiplex amplification of targets described in this instance but also in a more general

sense to provide methods which may be applied to reduce amplification costs and aid genotyping burden reduction in subsequent analyses.

Before association may be initiated however, a careful literature search must be performed to examine the disease pathogenesis and select suitable disease candidate loci or genes of appropriate size. The following section provides an introduction to NIDDM, the disease pathogenesis, symptoms and long term complications of the disease. The function of our three candidate genes are described and potential mechanismby which these may effect NIDDM development outlined.

1.2 Non-insulin-dependant-diabetes mellitus (NIDDM)

Non-insulin-dependant-diabetes mellitus (NIDDM), also known as type 2 diabetes is a chronic endocrine disorder characterised by insulin resistance, deficiency and hyperglycaemia [14]. Insulin is one of the key players in NIDDM development. Produced by pancreatic ß-cells this hormones primary functions are to facilitate cellular uptake of blood glucose and lipogenesis[1], facilitate increased amino acid transport into cells and reduce lipolysis[2]. It also stimulates growth, DNA synthesis, and cellular replication, activities which mirror those of insulin-like growth factors (IGFs) [14,15].

High blood glucose triggers pancreatic ß-cell uptake, leading to an elevation in the ATP/ADP ratio. This activates K^+ channel inhibition causing cell membranes to become depolarised and Ca^{2+} channels to become activated. The net result is electrically

[1] Lipogenesis is the processes whereby simple sugars are used for fatty acid synthesis and subsequent triglyceride synthesis.
[2] Lipolysis is the breakdown of fat stored in adipose cells.

stimulated insulin secretion and insulin action [16]. The order of insulin secretion should induce an appropriate cellular insulin receptor response inducing glucose uptake and the allowing blood glucose to return to normal. In cases of insulin resistance it is seen that although appropriate insulin secretion is induced, the analogous cellular response is not elicited. Accordingly, persistently high blood glucose, insufficient glycogen storage and hydrolysis of stored triglycerides (high blood tri-glyceride) are observed.

Exposure to high blood glucose such as that seen in hyperglycaemia stimulates beta cells to produce more insulin to effect the required blood glucose reduction, however over prolonged periods this high insulin output causes beta-cell apoptosis to increase and total beta-cell mass is reduced [17-19]. This is termed insulin sensitivity or deficiency and heralds a worrying phase in glucose homeostatic control. Insulin resistant profiles are often symptom-free and may frequently be reversed by diet and exercise, insulin deficient profiles often represent a more chronic phase and it is at this point that an individual will generally move from pre-diabetic to NIDDM classification [20-25].

NIDDM patients often present with manageable symptoms including thirst, weight loss, increased urination and tiredness, however with progression of the disease serious longer-term complications may arise. These predominantly develop from microvascular disease which can lead to renal complications including nephropathy, cardiovascular arteriosclerosis and retinopathy, an eye disease which can impair vision. Diabetic neuropathies are among the most frequent complication of long-term diabetes; with 60% to 70% of diabetics determined to have some form of nervous system damage which may manifest as pain, muscle weakness incontinence, oedema, neurogenic impotence and

paresthesia [26,27]. Less commonly seen is ketoacidosis, however the accumulation of ketoacids (a by-product of fat metabolism) can initiate severe effects; inducing diabetic coma and death in some instances [14].

Once this severe phase in NIDDM development is reached, the disease tends to be progressive and treatment and management of NIDDM and associated long-term complications costly. Approximately 5% of total NHS spend (3.75 billion) was consumed in management and treatment of NIDDM in 2007, a figure projected to rise in coming years[28]. However much of this cost is associated with chronic NIDDM profiles, as such determination of “at-risk” SNP profiles and implementation of individualised lifestyle management programs may act reduce the propensity for disease development, while careful monitoring of at-risk individuals may allow early-stage detection and avoidance of costly complications. The need for earlier disease monitoring was recently espoused by Douglas Smallwood, chief executive of Diabetes UK who said "many of the worst effects of diabetes can be avoided. We cannot afford to wait until people have heart attacks or have problems with their sight or kidneys before they get the care they need" [29].

1.3 NIDDM and Insulin-Like-Growth Factors

The genetic determinants of NIDDM development have been investigated extensively and a wide range of targets including insulin receptor-related receptor (INSRR), calpain-10 gene (CAPNIO), hepatic pyruvate kinase (PKLR), fatty acid binding protein (FABP2), peroxisome proliferators-activated receptor-gamma (PPARy), apolipoprotein A2 ATP-binding cassette, sub-family C (ABCC8) identified [30,31]. However further characterisation of these disease genes and an array of other potential targets must be investigated extensively if a comprehensive NIDDM disease profile is to be unravelled. The insulin-like growth factor (IGF) family represents a class of target which merits further investigation.

As the name suggests, insulin-like growth factor-1 (IGF1) is primarily involved in growth, however its structure and a number of its functions mirror that of insulin. In 1978, Rinderknecht and colleagues determined that human IGF1 with its single chain 70-amino acid polypeptide and 3 disulfide bridges shares significant homology with proinsulin [32]. Subsequent analysis by Ullrich *et al.,* also identified striking similarities between their corresponding receptors (IGF1-R and IR), both of whom are members of the transmembrane tyrosine kinase receptor subfamily [33]. In fact the similarities are such that IGF1 has also been found to bind the insulin receptor, activating identical signaling cascades as that of its insulin cognate, albeit at a significantly lower affinity ~ 0.01x [34].

Moxham *et al.,* first noted that tissues, which expressed both insulin and IGF1 receptors, also express a hybrid Insulin-IGF1 receptor [35,36]. These hybrids contain one IGF1 and one insulin alpha/beta heterodimer, however their function is more analogous to that of IGF1 receptors with respect to a higher affinity for IGF1 and reduced autophosphorylation. As such, increased hybrid receptor proportions have been seen to reduce insulin binding in these tissues [37]. This is particularly relevant as muscle falls within this category and is responsible for approximately 80% of whole-body glucose uptake. Thus, even marginally altered hybrid receptor proportions may induce significant effects on glucose management and contribute to insulin sensitivity [36].

The role of IGF genes and relationship to known NIDDM contributors such as insulin and growth hormone (GH) began to be investigated in the late seventies. Chronic elevation of GH has been shown to induce insulin sensitivity and as such, factors effecting GH expression are worthy disease candidates [38]. The correlation between circulating IGF1 concentrations and GH expression is striking; a recent study by Haluzik *et al.,* found down-regulation of IGF1 by 65%, 75% and 85% corresponded with no change, a 4X increase and a 10X increase respectively in GH production in engineered mouse models [36]. This relationship between IGF1, GH and NIDDM was further validated by Yakar *et al.,* who showed treatment of liver IGF1-deficient (LID) mice with exogenous IGF1 resulted in inhibited GH secretion and improved insulin sensitivity [39,40].

However serum IGF1 is not maintained as a singular structure; a large proportion of IGF1 is complexed in the blood with other stabilising agents, which facilitate increased half-life and access to target tissues. In fact just 5% of blood IGF1 is present in its unbound form. Around 15% is composed of binary IGF1/IGF-binding protein complexes while the majority, ~80%, is found in a ternary complex of IGF1/IGFBP3 acid labile subunit format. This ternary structure crosses the capillary barrier poorly and thus acts as a stable reservoir of circulating IGF1 [36]. IGF1 is freed when required by proteolytic cleavage of IGFBP3 and interaction with proteoglycans [41]. As unbound IGF1 has a half life of <10 minutes compared to 12-15hrs for its ternary structure cognate, it follows that inadequate IGFBP3 activity may contribute to NIDDM by poor ternary complex formation with IGF1 and insufficient IGF1 reservoir maintenance [42]. The relevance of IGFBP3-induced stability was further highlighted by Bang *et al.,* who noted that increased serine protease activity (known to degrade IGFBP3) co-segregates with NIDDM phenotypes [43].

Recombinant DNA technology has also been utilised with good effect to investigate the use of IGF1 and its binding proteins for potential NIDDM treatment. Administration of free IGF1 was found to significantly reduce insulin resistance in NIDDM patients [44]. However administration of complexed IGF1/IGFBP3 has been used with even greater effect for treatment of insulin dependant diabetes mellitus (IDDM). This binary structure was found not only to reduce insulin requirements and lower serum glucose however, but a marked decrease in serious side effects were also noted [45,46].

In addition to its IGF1-dependent activities, IGFBP3 has also been found to display IGF1-independent functions affecting growth and apoptosis and which may modulate disease progression and longer term NIDDM-associated complications [47]. Association between IGFBP3 and retinal neovascularisation (a leading cause of diabeteic retinopathy) was recently highlighted by Lofqvist *et al.,* who identified correlation between increased IGFBP3 levels and reduced retinal neovascularisation in oxygen-induced retinopic mice. Results indicate that IGFBP3 acts (independently of IGF1), as a progenitor cell chemoattractant, reducing oxygen-induced vessel loss and promoting vascular regrowth [48]. IGFBP3 is thus an attractive candidate treatment for prevention of diabetic retinopathy, the leading cause of blindness in persons of less than 75 years [49].

Table 1: Effect of candidate gene IGF1 and binding proteins IGFBP1 and IGFBP3 on blood glucose and NIDDM development. Increased IGFBP1 and reduced IGFBP3; reduce IGF1 bioavailability and half-life resulting in increased blood glucose. Hyperglycaemia induces increased propensity towards insulin sensitivity and NIDDM development [36,50].

↑ IGFBP1 Or ↓ IGFBP3	↓ IGF1	↑ Blood Glucose	↑Insulin sensitivity, ↑ NIDDM

In total 6 IGF binding proteins exist. IGFBP1 is a particularly strong NIDDM gene candidate as it is the only known acute regulator of IGF1 bioavailability [42]. High IGFBP1 expression has been found to correlate with reduced IGF1 activity, resulting in poor IGF1-stimulated growth, differentiation and hypoglycaemic control. This reduced IGF1 availability has been found to induce increased insulin resistance and glucose intolerance [51]. With regard to NIDDM phenotype induction; the mechanism of IGFBP1 action may be postulated from a study by Uekiet *et al.* Using mouse knockouts lacking

IGF1 receptors, Uekiet *et al.,* found poor IGF1 uptake to correlate with reduced glucose-stimulated secretion of insulin without beta cell mass attrition [52]. It may thus be postulated that a similar effect may be induced by over-expression of IGFBP1 and resultant IGF1 unavailability.

Thus clear supportive evidence exists regarding the roll of IGF1 in glucose homeostasis and multiple proposed modes of action regarding how IGF binding proteins one and three may contribute to the NIDDM disease profile. IGF1, IGFBP1 and IGFBP3 gene targets were therefore selected for further association-based SNP prioritisation.

1.4.1 Aims

The aim of this work is to use haplotype based tagSNP selection and a systematic *in silico*-based analysis approach to design a multiplex compatible PCR primer and SAT probe set facilitating maximum variation capture with minimum tests across candidate genes IGF1, IGFBP1 and IGFBP3. This will be applied though development of a number of robust, high-efficiency, high-specificity multiplex PCR constructs for amplification of multiple targets to demonstrate the applicability of these target types to suspension array genotyping for non-insulin-dependant diabetes mellitus.

1.4.2 Objectives

- To construct haplotypes for a given population and perform tagSNP selection which captures maximum variation across candidate genes IGF1, IGFBP1 and IGFBP3.
- To perform multiplex primer and probe design, utilising *in silico* and manual analysis for evaluation and selection of a high specificity primer and probe sets in a manner compatible with downstream multiplex PCR and SAT analysis.
- To perform extensive PCR optimisation for the construction of a number of robust, well characterised, high dimensionality multiplex PCR sets.
- To demonstrate the applicability of multiplex PCR to suspension array facilitated allele discrimination using a model probe pair and validate allele designation via dot blotting.

1.4.3 Thesis overview

Chapter 1: Introduction

Polygenic disease is introduced. Confounders associated with polygenic disease discovery are described and innovations to reduce problem complexity and confounding with regard to study design and experimental performance are discussed. NIDDM is introduced and hypotheses regarding candidate IGF1, IGFBP1 and IGFBP3 gene participation in NIDDM predisposition reviewed.

Chapter 2: Haplotype based TagSNP Selection and Functional Impact Prediction

Linkage and association-based disease discovery techniques are described with particular emphasis on the efficiency and accuracy of haplotype and functional impact-based polymorphism prioritisation methods. Haplotype-based design considerations, parameter selection and thresholds designation are investigated in a bid to improve the likelihood of true quantitative trait loci (QTL) discovery. Haplotype-based prioritisation techniques are applied to candidate IGF1, IGFBP1 and IGFBP3 genes for optimal marker selection in Caucasian populations and captured SNPs are analysed to predict their putative functional impact.

Chapter 3: PCR Primer and probe Design

The difficulties associated with increased dimensionality PCR amplification and suspension array genotyping are introduced with particular emphasis on primer and probe design. Oligonucleotide design features which may be considered to increase the probability of successful multiplex amplification and genotyping are discussed and a systematic method using both manual and *in silico* evaluation applied to candidate tagSNPs (selected in chapter 2) for construction of a high-specificity primer and probe set with a low putative propensity toward aberrant functionality. Final primer and probe set profiles determined using this method are described and critiqued with respect to their proposed multiplex application.

Chapter 4: Multiplex PCR Amplification

This chapter sets out to evaluate primer pairs (designed in chapter 3) in terms of true experimental functionality. Multiplex-critical primer-pair parameters are investigated extensively using singleplex end-point and real-time PCR amplification and results discussed with respect to theoretically derived profiles. Results of singleplex experimental amplification are used to inform multiplex set construction and optimisation and the effect of increased dimensionality formats on amplification efficiency discussed. This chapter also provides an overview of reaction components adjuvants which may be used to ameliorate problematic PCR amplification and a number of these techniques are implemented for amplification of repeat region targets in this instance.

Chapter 5: Suspension Array SNP Genotyping

This chapter aims to introduce the area of flow cytometry facilitated suspension array; comparing it to other planar array and SNP genotyping methods currently available. Linear probes which facilitate genotyping of PCR amplicons are designed using manual and *in silico* methods as previously described (chapter 3). A model probe pair is used in a proof of concept study for SNP genotyping of target amplicon sequences. Probe coupling and target hybridisation efficiency using single stranded and multiplex targets are described. Allele discrimination designation is validated via sequencing of target amplicons.

Chapter 6: Final Discussion, Conclusion and Future Work

The final chapter considers and discusses the work presented in this dissertation. Results described are discussed with respect of the aims and objectives of the project and with a view to its wider potential applicability for disease association. A conclusion regarding the work is drawn and further work pertaining to this project is discussed.

Chapter 2: Haplotype Construction and TagSNP Selection

2.1 Introduction

Although animal and cell line models have been instrumental in deciphering the complex pathways that underlie NIDDM, they are not without their limitations. Ethical constraints, limited genomic and pathway homology, disease heterogeneity and increased costs associated with use of highly syntenic models mean that genomic investigation, such as those afforded by forward genomic approaches presents an attractive alternative. Forward genomics aims to allow the genetic determination of observable phenotypic variation. Two broad approaches may be taken; linkage (genome scan) or association analysis (candidate gene) [53].

This chapter serves as an introduction to forward genomic disease discovery techniques; linkage and association analysis are reviewed with particular emphasis on applicability to polygenic disease discovery. Association-based design considerations and parameter and thresholds selection are investigated in a bid to improve the likelihood of true quantitative trait loci (QTL) discovery. Association-based techniques are then applied to candidate IGF1, IGFBP1 and IGFBP3 genes for optimal marker selection in the target population and SNPs analysed to predict their putative functional impact.

2.1.1 Linkage Analysis

Linkage analyses aim to follow meiotic events through pedigrees to identify increased co-segregation of alleles sharing distribution with disease traits at a rate greater than that which would be expected by the laws of independent assortment. These loci should be linked with the true disease susceptibility loci but may not contribute to the disease profile themselves.

Pedigrees are used to minimise the number of markers required for adequate inheritance mapping. As recombination rates within families are low, identity by descent (IBD) haplotypes tend to be inherited in large megabase blocks, rather than the kilobase sized IBDs displayed by more distant population structures. Recombination frequency is predominantly a function of physical distance on a single chromosome and as such, may be exploited by use of mapping functions to convert recombination frequencies between adjacent loci to genetic map distances [54]. In this way genetic linkage maps, detailing the likelihood of recombination events between ordered sets of evenly spaced markers, have been constructed. The principle of identity by descent mapping is outlined in Figure 1.

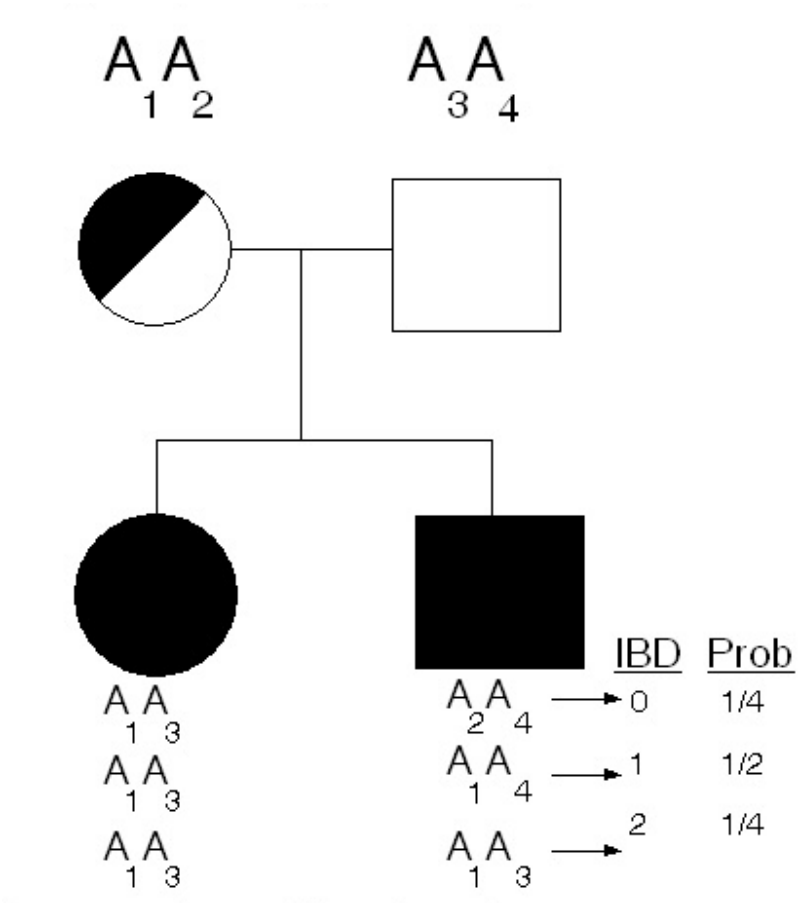


Figure 1: Principle of IBD Mapping (Identical by Descent Mapping). Two genes/alleles are identical by descent if they are exact copies of the ancestral gene/allele. Two siblings can share 0, 1 or 2 parental marker alleles identical by descent at any locus with respective probabilities 0.25, 0.5 and 0.25 under random segregation. Male and female symbols are represented squares and circles respectively [55].

Low density microsatellites maps (marker intervals of ~10cM[3]) have traditionally been used for this purpose, however high density SNP maps have now been developed with good effect [56]. Although individually less informative than microsatellites, SNPs are distributed at a much higher rate throughout the genome, allowing map intervals of less than 1cM to be set. This approach has proven to be more powerful in terms of linkage localisation and generation of significant linkage scores [57,58]. In addition genotyping error rates tend to be lower for SNPs than for microsatellites and when parental genotype information is considered; lower type I error rates have also been found to be produced [59,60].

Either model based (parametric) or model free (non-parametric) methods may be implemented, although in truth parametric models are rarely useful for true polygenic disease discovery. Disease model specification is required for parametric linkage (PL); however polygenic diseases are constituted by casual genes which display small individual effects. These genes require multiple contributory loci, and complex interactions in order for their effect to be felt, making the polygenic disease profile difficult to reconstruct. Additionally polygenic disease onset is generally later; making construction of large multigenerational pedigrees difficult. Conversely nonparametric (NPL) techniques do not require model specification. Small pedigree generations and an array of pedigree relationship types including affected sib-pair, parent-child pair and cousin pair relationships may therefore be used.

[3] A centimorgan (cM) is a unit that describes a recombination frequency of 1% and is approximately 1,000kb in the human genome.

Two-point or (more commonly) multipoint linkage analysis may be used to evaluate NPL. Two-point methods use logarithm of odds (LOD) scores to identify regions of increased co-segregation between single markers and disease traits. However because single marker alleles do not always define the inheritance pattern (i.e. grand paternal/maternal founder member who passed the allele) power can be lacking.

Multipoint analysis allows information from multiple markers to be used to infer IBDs and calculate linkage scores [61]. Calculation of full multipoint IBD distribution for pedigrees is a computationally intensive exercise however; and methods that calculate exact maximum likelihoods (Elston-Stewart, Lander-Green) are limited in terms of the number of loci and/or individuals that may be analysed concurrently. The Markov-Chain Monte-Carlo (MCMC) method bypasses this problem by instead estimating maximum likelihoods and as such, MCMC can process large pedigree loads and multiple markers in excess of 1000 loci and individuals simultaneously. The computational efficiency afforded by MCMC also facilitates analysis of epistasis and genetic heterogeneity via a number of two-locus linkage enrichment models [62,63].

Although initially unconsidered in single-locus models, analysis of co-segregation between casual genes on alternate chromosomes can serve to strengthen linkage scores. Several two-locus models have been devised using various approaches including, LOD score calculation (with an assumption of a two-locus disease model) and a two-stage approach which assesses correlation between unlinked regions prior to estimation of joint susceptibility induction. The latter model is particularly useful for enrichment of samples where genetic heterogeneity is suspected [64]. These methods may be implemented for

discrete or quantitative traits and are readily available in programs such as Loki[4] and MORGAN[5] . With regard to quantitative traits however, both are restrictive in that only additive multi-locus models may be applied [62]. An overview of linkage analysis from pedigree marker analysis through to SNP association is outlined in Figure 2.

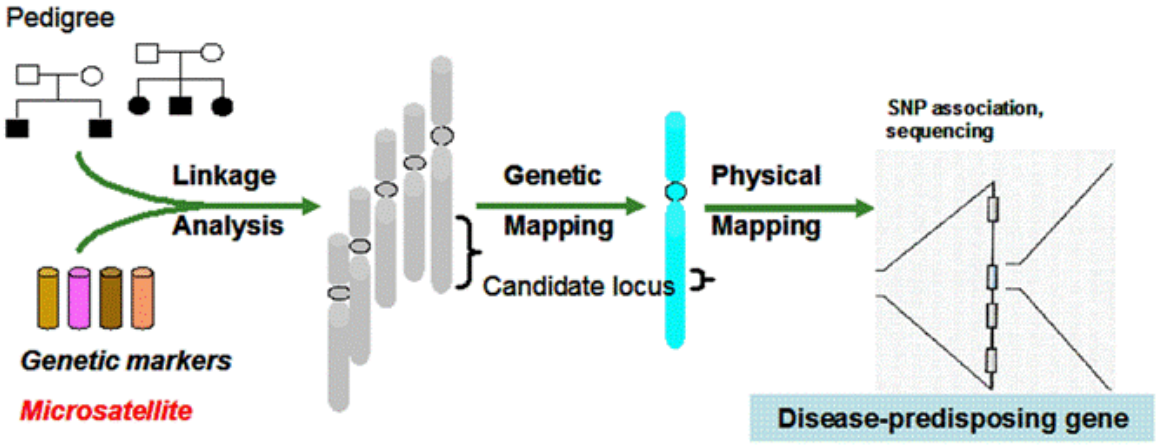


Figure 2: Linkage Analysis; Markers spanning the genome are typed and used to map IBD haplotypes within pedigrees. Excess IBD haplotype sharing (greater than that expected by the law of independent assortment) within cohorts groups indicate regions of potential disease linkage. Further analysis of Linkage region may yield functional polymorphisms.

Linkage analyses have been used to convincingly identify many disease loci in multiple populations, including ten NIDDM linked regions on chromosomes 1q25.3, 2q37.3, 3q28, 3p24.1, 6q22, 8p23, 10q26.13, 12q24.31 18p11.22 and 20q13.1 with consequential genome-wide scores, in excess of that which is considered significant (LOD >3.0) [31].

[4] Loki: http://www.stat.washington.edu/thompson/Genepi/Loki.shtml

[5] MORGAN: http://www.stat.washington.edu/thompson/Genepi/MORGAN/Morgan.shtml

However Linkage analysis also has a number of drawbacks, rendering the approach incompatible with many aspects of the polygenic disease profile.

One of the main difficulties is that linkage analysis provides relatively low statistical power for detecting QTLs with small to modest effects, i.e. given strong linkage ($\theta = 0.05$) Sham *et al*., determined that a sample size of 80,620 is required to detect QTLs conferring 5% trait variance. In practical terms this renders linkage analyses aiming to capture variance of <10% infeasible and given the putative nature of polygenic disease, a large proportion of QTLs conferring small to moderate risk may be overlooked if linkage alone is employed [65]. The use of pedigrees for analysis of late onset diseases also presents particular problems in terms of sample collection in Western Europe where families have declined in size dramatically over the past 60 years [66]. Additionally, although use of this cohort type facilitates large IBD haplotype coverage with minimal marker requirements, it also precludes fine mapping and identification of individual linked genes or casual polymorphism alleles. As such population-based association analyses have been developed in an attempt to alleviate these problems and make forward genomics amenable to a wider range of polygenic disease models. The following section covers disease association; describing and evaluating both functional prediction and linkage disequilibrium-based approaches.

2.1.2 Association Analysis

Association analysis measures the relative frequency with which marker alleles co-segregate with disease traits within a cohort. However unlike linkage analysis, association does not exploit linkage but rather linkage disequilibrium. Linkage Disequilibrium (LD) describes non-random association between two or more alleles across the genome. When two or more regions are in strong LD they are usually inherited together without recombination. Consequently, association is not limited to assessment of correlation between markers and disease on individual chromosomes and epistatic interactions are more readily modelled by association approaches. This method is generally used, *a posteriori*, to finely map regions or individual SNPs of interest highlighted by genome-wide linkage analysis, expression analysis, functional studies, or functional prediction methods and aims to identify susceptibility alleles or markers in close linkage disequilibrium with true at-risk polymorphisms [67,68].

Disease discovery can be time consuming as multiple markers and many individuals must be tested in order to collate a comprehensive set of susceptibility alleles, as such establishing priorities in the selection of SNPs based on functional prediction or linkage disequilibrium is an excellent way to accelerate the process and reduce cost.

2.1.2.1 Functional Candidate Based Prioritisation

The functional candidate approach prioritises individual polymorphisms based on their putative capacity for functional disruption. Selection may be based on the genomic environment in which they reside and/or on the extent of their presumed capacity for functional disruption. An array of computer programs and algorithms have been developed to simplify the selection process; sequence and 3D-structure homology or RNA stability modelling may be implemented for SNPs located in coding regions. While splice / branch site motif analysis may be used for functional prediction of SNPs located in non-coding regions.

The following section gives an overview of the methods that may be employed for *in silico* functional impact prediction and Table 2 provides a list of some of the software programs available for this type of application. A number of models are currently used to predict the functional impact of coding SNPs on protein function: Sequence homology uses retention of amino acids in homologous proteins to predict tolerance of residue change while 3D structure analysis exploits use of known protein structures to model the effect of residue change on electrostatic interaction, catalysis, ligand binding and hydrophobic disruption of the protein core. Research suggests that use of multiple analysis methods including that of 3D structure modelling result in highly accurate prediction, however just 60% of proteins have been modelled to date making modelling of this nature inaccessible for many protein products and isoforms. This multiple analysis approach used by Sunyaev *et al.,* for human transthyretin functional impact prediction is outlined in Figure 3.

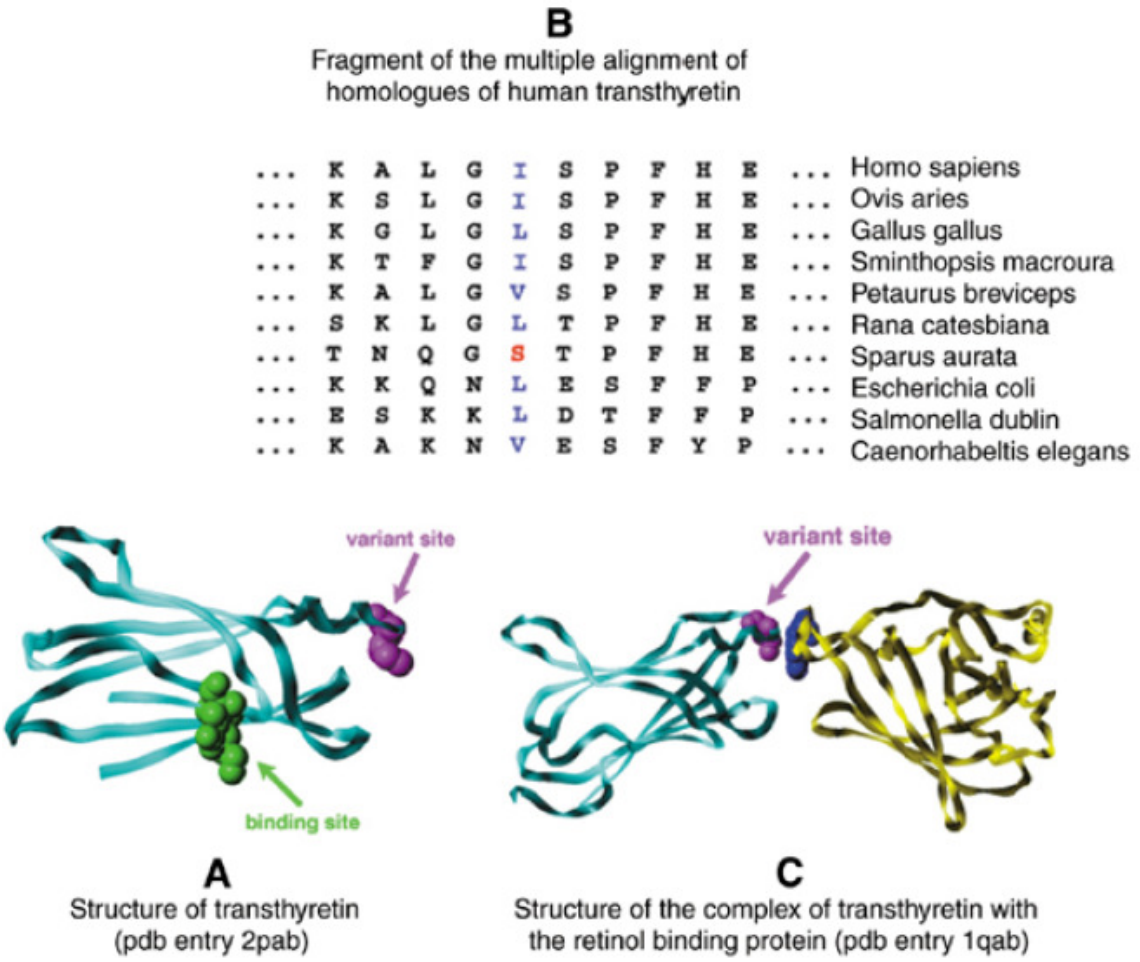


Figure 3: Multiple prediction mechanisms can help improve the accuracy of functional impact prediction. Comparative residue analysis of Ile-Ser replacement in transthyretin fails to predict a deleterious impact as this residue can also be seen in the Sparus aurata homolog. Although the substitution results in a hydrophobic to hydrophilic residue change on the molecule surface, structural analysis also identifies this substitution as neutral, as its location is quite a distance from the binding site. However, complex formation analysis predicts the residue to play a role in substrate complex formation and a deleterious prediction was duly made [69].

RNA stability modelling is currently used to predict the impact of larger polymorphisms on single stranded RNA molecular stability. Using this method conformational energy searching is carried out to find the energetically preferred conformations of a molecule.

Although local instability may be necessary to facilitate certain mRNA activities (i.e. catalysis and binding) alleles which produce the lowest free energy measurements are generally more stable and should therefore incur higher activity [70,71]. The extent of free energy difference between RNA molecules and resultant RNA folding changes can be used to predict the putative impact which may be imparted (Figure 4). Structural changes of this nature can be stark impacting complex formation between mRNA, tRNA and ribosomes and having knock-on effects on the protein expression.

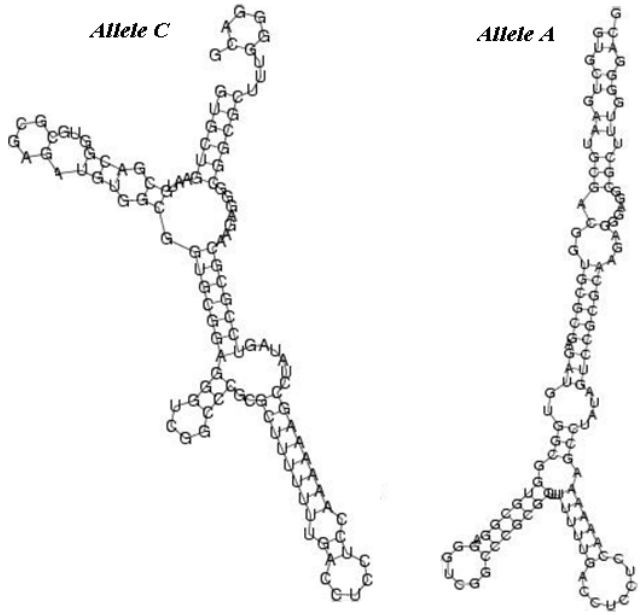


Figure 4: RNA Analyzer; local minima are sought to model RNA structure. A single allele change can significantly effect the theoretical structure produced and may have real life implications for RNA structure / function. The example featured displays a modelled 114 base sequence with alternate alleles C or A at position 99. A clear difference may be observed between structures, this was also reflected in energy scores [70].

Non-coding regions are segments of DNA that do not encode functional proteins or RNA. Such regions do however; often contain conserved regions involved in control of expression, or that play a role in structural stability of the genome. Promoters work in

unison with other regulatory regions such as enhancers, silencers and insulators to direct the level of transcription of a given gene. SNPs located in transcription factor binding sites (TFBS), triplex-forming oligonucleotide target sequences (TTSs), catalytic regions, poly-A tails and other conserved motifs are likely to direct more significant change than those located in non-conserved regions.

Clear donor and acceptor splice signals are important for defining splicing boundaries and recognition of exons, while exon splice silencers (ESS), exon splice enhancers (ESEs) and intronic splice silencers (ISS) are known to direct selection and expression of alternate transcripts.

The success of *in silico* prediction strategies depends heavily on how well putative functional variants are identified however. If prediction and thus prioritisation is poor, the proportion of true functional polymorphisms included may be too small to facilitate standard association detection. While some prediction methods and programs display encouraging detection rates (i.e. PolyPhen true positive non-synonymous SNP prediction of 82%), others, such as RNA stability/structure analysis, are little more than speculative (at least in terms of SNP analysis) [72,73] (private correspondence, Zucker, 2007). In addition, the limited knowledge of promoter and other non-coding regions available means that accurate prediction of SNP effects in these environments is more challenging [67].

Table 2: *In silico* Functional Impact Prediction Software Programs. A number of analysis implementations are listed, along with the general category of prediction facilitated.

Functional prediction Category	**Software program**	**Software access**
Protein catalytic sites, structure and binding	SIFT (Sorting Intolerant From Tolerant) PolyPhen (*Poly*morphism *Phen*otyping) SNPs3D	http://blocks.fhcrc.org/sift/SIFT.html [74] http://genetics.bwh.harvard.edu/pph/ [75] http://www.snps3d.org/ [60]
Post translational modification, protein aggregation and amyloidosis	SNPeffect	http://snpeffect.vib.be/ [76]
mRNA stability modelling and regulatory motifs	RNA Analyzer GeneBee	http://wb2x01.biozentrum.uni-wuerzburg.de/ [60] http://www.genebee.msu.su/genebee.html [77]
Splice site recognition and regulation	Automated Splice Site Analysis	https://splice.uwo.ca/ [78]
Comprehensive genomic analysis (Protein, Post translational modification, Splice site, intron/ exon boundary and motif analysis)	PupasView	http://www.pupasnp.org [79]

Initially, non-synonymous coding SNPs were believed to be responsible for the majority of polygenic disease predisposition and as such, candidate SNP selection was weighted heavily towards this premise. One of the most frequently cited NIDDM susceptibility polymorphisms; the missense alanine pro12Ala allele, was initially highlighted as the at-risk allele. However more recent association studies have refuted this claim and an investigation by Wei *et al*., using alternate ethnic cohorts indicate that this SNP is more likely to be in linkage disequilibrium (LD) with the true at-risk allele [80].

This pattern of weak and confounded association is repeated in multiple association studies across candidate genes IGF1, IGFBP1 and IGFBP3 chosen for this study, where markers have been predominantly selected via functional impact prediction. Two of the most heavily studied NIDDM candidates across this region are; IGF1 non-synonymous coding SNP Gly972Ar (rs1801278) and promoter constituent (CA)n repeat polymorphism 192-bp. Vaessen *et al*., initially reported evidence of NIDDM association with IGF1 cytosine-adenosine non-192-bp repeat promoter polymorphism. This allele was found to confer an 18% reduction in IGF1 expression and an increased relative NIDDM risk of 1.7, however a subsequent analysis by Frayling *et al*., using a UK Caucasian cohort did not support these findings [81,82]. A third study by Rietveld *et al*., did not support the original findings either, and instead found correlation between homozygous carriers of the 192 bp allele and increased GH-driven age related decline in total circulating IGF-I [83]. This indicates that the polymorphism itself is unlikely to be functional, or that heterozygosity in study design contributed to conflicting outcomes.

Tentative association between NIDDM and IGF1 nonsynonymous SNP Gly972Arg was also identified by Florez *et al*., although limited power meant subsequent investigations of this SNP again produced conflicting results. A meta-analysis undertaken by Jellema *et al*., found carriers of the 972Arg variant to induce a substantial 25% increased risk of NIDDM by comparison with non-carriers [84]. However two subsequent large-scale population studies of Caucasian cohorts (n=9,000 and n=1,467 respectively) found this substitution to have little appreciable effect on common NIDDM predisposition [85,86].

Studies such as these highlight the potential pitfalls of association; large sample sizes, use of homozygous populations and meta-analyses do not necessarily render association clear-cut. The process of association is a complex one; confounding can occur as a result of multiple factors irrespective of the markers chosen or manner in which markers were selected, however use of a more comprehensive haplotype-based approach may offer a method to further minimise confounding caused by linkage disequilibrium and incomplete gene coverage. The following section describes haplotype and linkage disequilibrium-based marker prioritisation.

2.1.2.2 Haplotype Based Prioritisation

An alternate marker haplotype-based prioritisation approach uses linkage disequilibrium to guide selection of markers representing a high proportion of common variation across candidate regions.

As previously discussed, increased physical distance generally results in increased recombination and lower linkage disequilibrium along the length of the whole chromosome. This also holds true for smaller, minimally recombinant haplotypes, with alleles positioned at the edges of haplotype blocks experiencing greater recombination than haplotype-central alleles (Figure 5). As such individual SNPs may be used to capture the maximum amount of variation across the haplotype and predict which alleles are likely to be present. These representative SNPs are termed tagging SNPs, tagSNPs or haplotype-tagged SNPs and have been found to be highly effective for disease association of common alleles [6].

While such an approach has clear application for high frequency alleles, Prichard *et al*., additionally postulates that it may offer advantages for SNPs subscribing to the rare allele common disease hypothesis [87-89]. This is based on observations by Patterson *et al.,* who noted that rare SNPs often appear within long-established haplotypes; as such, it is thought effects may be more detectable within haplotype-tagged studies than with single variation approaches [67].

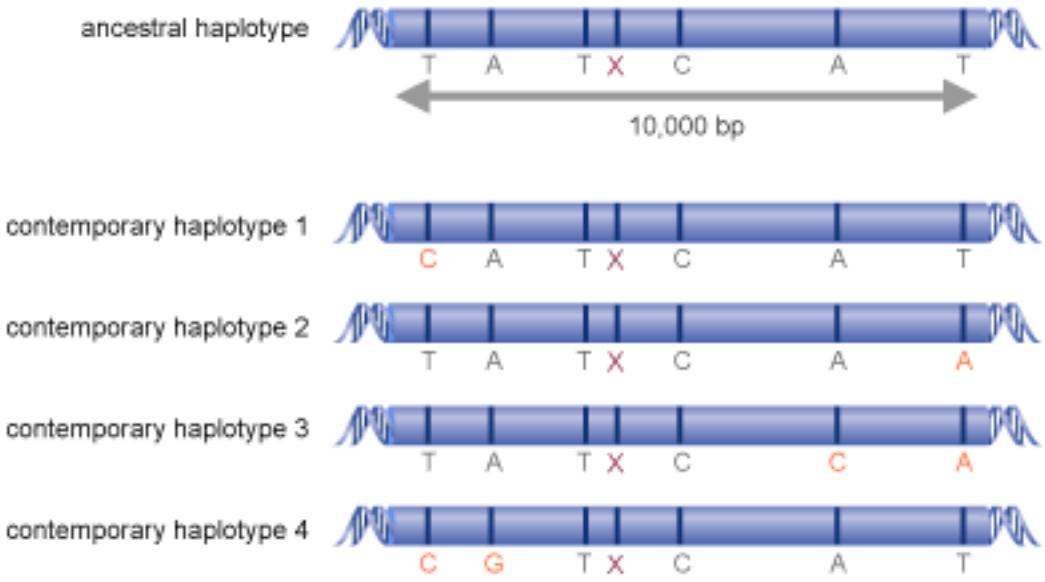


Figure 5: Linkage Disequilibrium within a Haplotype Block over Multiple Generations. A new mutation (X) arises in the ancestral haplotype signature TATCAT that causes disease. Within the haplotype block recombination occurs more frequently with increasing physical distance. As such the most highly conserved haplotype signature in patients carrying the disease allele are central around the casual allele.

Power for detection may also be enhanced for haplotypes containing multiple contributory SNPs subscribing to an additive disease model, as linkage disequilibrium allows the impact of all tag and capture SNPs to be felt.

However while Haplotype-based tagSNP association has a higher practical power capacity than functional prediction, approaches are known to suffer from confounder susceptibility which can render true association elusive [88]. The validity of results obtained can be heavily influenced by selected algorithms, thresholds and the study design implemented; therefore, it is essential that analysis options are evaluated and selected carefully before any association study is initiated [90].

The following section describes potential confounders, design features and options that may be considered prior to haplotype- / LD-based association analysis to minimise confounding and maximise the probability of true association. Haplotype inference, LD calculation and LD-based tagSNP selection / paring approaches are all discussed.

2.1.3 Haplotype Inference and Linkage Disequilibrium Calculation

Somatic cell chromosomes exist in pairs of homologues, as such it is necessary to infer phase (which polymorphisms are derived from each homologous chromosome) when estimating haplotypes and their associated frequencies.

A number of techniques have been developed for direct haplotype observation including single-molecule dilution, long-range allele specific PCR, pyrosequencing, intracellular ligation, rolling-circle amplification, carbon nanotube probing, diploid-to-haploid conversion and clone-based systematic haplotyping. However these techniques are often labour intensive, costly and hampered by technical difficulties which make large scale application difficult [91-93].

Haplotypes may be directly inferred from tri-generational pedigree data although most polygenic association studies use less complex pedigree structures or data from unrelated individuals that cannot accommodate this approach. As such, a number of alternative haplotype estimation-based approaches have been developed including Clark's algorithm, expectation-maximisation (EM), coalescence and Bayesian-derived methods [93,94].

These algorithms can predict multiple haplotype frequencies for un-phased individuals, however phase uncertainty remains a feature of all and this uncertainty influences subsequent haplotype construction and LD calculation.

For un-phased genotypes the haplotypes inferred by these programs are described with a certain probability, however the linkage disequilibrium calculations which follow generally treat the most probable haplotypes as true and subsequent calculations are based on these assumptions. Kulle *et al*., described a method to derive LD measures from relative haplotype frequencies produced by EM or Bayesian approaches using weighted or joint statistical models, essentially incorporating haplotype uncertainty into LD calculation. While this may benefit Mendelian disease discovery however, Lu *et al*., determined that this is likely to be of little benefit to complex disease analysis due to the increased model complexity and relatively weak gains imparted [95,96].

A host of LD measurements exist, with disequilibrium statistics (D') and pair-wise correlation coefficients (r^2) being the most common. D' displays indicator-like behaviour for missing haplotypes and tends to enrich for rarer alleles, while r^2 is more favourable for common alleles. Additionally, r^2 is inversely related to power (i.e. sample size) and as such, has become popular for tagging and disease association studies (see Figure 6)

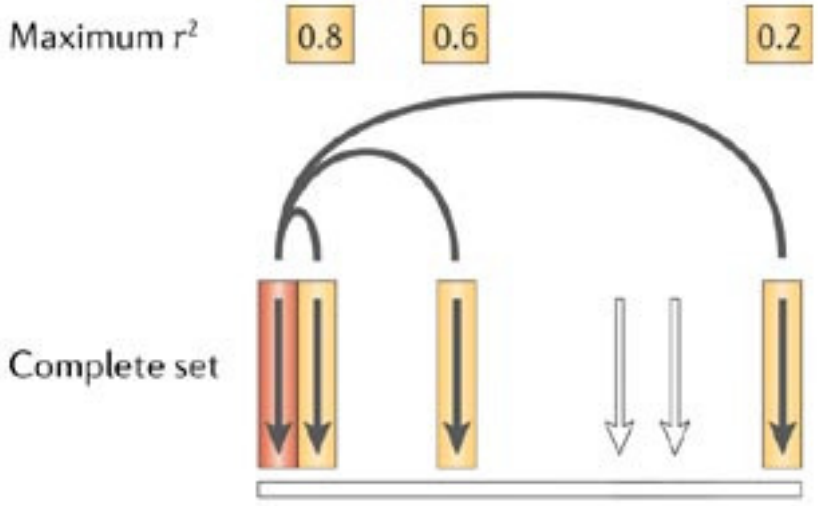


Figure 6: Association analysis sample size calculation. Maximum r^2 can be used to calculate the sample size requirements for indirect association. Both the frequency of marker and capture alleles and the strength of LD between markers and capture SNPs influence sample size requirements. Higher minor allele frequencies and LD (maximum r^2) between tag and capture SNPs results in reduced sample size requirements [97].

2.1.4 LD-based tagSNP Selection Methods

Prioritisation, and thus reduction, of tagSNP burden can be accomplished using haplotype block-based or block-free methods [98]. Block-based methods such as those developed by Johnson *et al.,* Patil *et al.,* and Zhang *et al.,* are based on reconstruction of LD using haplotype block models in candidate regions (see Figure 7 for graphical representation). Resultant tagSNPs aim to describe variation across individual blocks using a subset of non-recombinant SNPs [99]. These methods work best when small numbers of common haplotypes exist within the population sample and require full phase information and block boundary identification. The definition of a haplotype block partition has been much discussed with three established classes (diversity, LD and recombination-based methods) found by Ding *et al*., to produce divergent haplotype block partition and alternately sized

haplotype tagged SNP subsets. The effect of haplotype block definitions on subsequent associative power has yet to be resolved; however these findings underline the inconsistencies that may impact power to detect true association when implementing block-based SNP prioritisation [100].

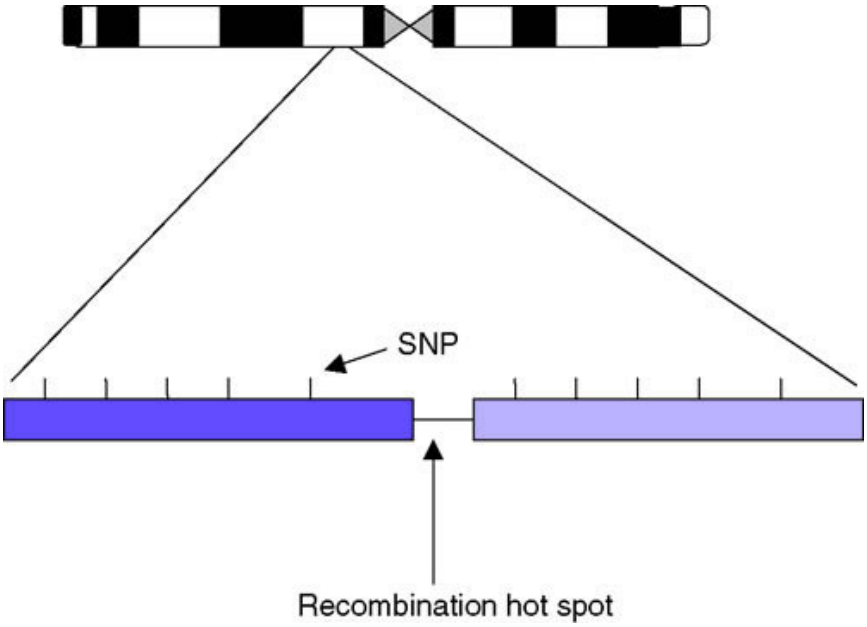


Figure 7 Haplotype block-based tagging. The genome is organised into 10- to 20-kb haplotype blocks that are in strong LD and are transmitted more or less intact from generation to generation. Separating these are "recombination hot spots" where recombination is most likely to occur. As SNPs within a block tend to be in LD with each other; single tagSNPs can be used to predict the genotype of multiple markers within the block [101].

Haplotype block-free methods such as those described by Carlson *et al.,* Halldorsson *et al.,* Halperin *et al.,* He *et al.,* and De Bakker *et al.,* aim to capture maximum variation across the candidate region as described by linkage disequilibrium, Recombination is not

considered however, and block boundaries are not defined [6,7,99,102]. Consequently, uncertainty associated with block definition is circumvented and correlation between alternate haplotype blocks may be exploited with reduced tagSNP burden [99]. This can increase associative power for haplotypes which subscribe to an additive effects model across a broader genotypic range while potential small sample size confounders may be limited by setting LD recognition windows to a physical distance suggestive of true LD [99,103].

Other additional strategies may be employed to reduce tagSNP burden, include r^2 correlation threshold reduction and use of best "N" methods. At present most tagSNP selection is based on pair-wise LD, capturing minimal sets of markers based on pair-wise r^2 correlation between adjacent SNPs. SNPs are in perfect LD if $r^2 = 1$ however slightly reducing this correlation threshold, can dramatically decrease the tagSNPs requirement while maintaining high relative power to detect association for common casual alleles. De Bakker *et al*., found an r^2 threshold of > 0.8 reduced tagSNP requirements by 55% (in Central European ancestry Utah populations) while retaining relative associative power of 96%. Further reduction of set thresholds results in rapid atrophication of relative power, and as such an r^2 threshold of > 0.8 is generally accepted as a suitable concession [7]

De Bakker further reduced tagSNP burden by development of a block-free multi-marker "best N" approach. This works by ranking potential tagSNPs with regard to the number of SNPs captured and has been found to be more efficient than the r^2 threshold reduction technique when complete reference panels are used. However as "best N" discards putative

self-tagging SNPs, incomplete LD may result is rejection of unlisted SNPs, a strategy likely to impact preferentially upon rarer alleles [7].

Although multi-marker approaches produce fewer tagSNPs they are more likely to exhibit lower associative power than pair-wise methods, as such, care must be taken to avoid over paring tagSNP lists [104]. The software program Haploview (http://www.broad.mit.edu/haploview/haploview-downloads) developed by Barrett *et al.,* at the Broad institute MIT has implemented a method which combines the simplicity of pairwise methods with the potential efficiency of multimarker approaches. Overfitting is avoided by using only those multiallelic combinations in which the alleles are themselves in strong LD [105].

In this study we use Haploview to infer phase (section 2.1.3 Haplotype Inference and Linkage Disequilibrium Calculation) construct haplotypes (section 2.1.2.2 Haplotype Based Prioritisation) and calculate linkage disequilibrium (section 2.1.2 Association Analysis) from unphased non-pedigree population data. TagSNPs are selected using Haploview's aggressive mode which implements a multi-marker haplotype-free tagging approach (section 2.1.4 LD-based tagSNP Selection Methods) to reduce tagSNP burden and increase associative power. TagSNP burden is further pared using r^2 correlation threshold adjustment (section 2.1.4 LD-based tagSNP Selection Methods). The following work sets out to address the first thesis objective "*To construct haplotypes for a given population and perform tagSNP selection which captures maximum variation across candidate genes IGF1, IGFBP1 and IGFBP3*".

2.2 Materials and methods

2.2.1 Genotype Data Collation and Amendment

High quality reference panels collated by NHLBI Seattle SNP's PGA (programs for genomic applications) and NIEHS's EGP (environmental genome project) were used to access high quality genotyping data for candidate genes. Where possible data included was limited to populations of European Caucasian descent (i.e. for IGF1 and IGFBP3), where such selections were unavailable (i.e. for IGFBP1); data was treated to remove putative African derived genotypes which tend to display increased genotype / haplotype diversity relative to Caucasian profiles.

Genomic data for both IGF1 and IGFBPP3 was accessed through GVS (Genome variation Server) website at http://gvs.gs.washington.edu/GVS/. *PGA* (programs for genomic applications) data was selected for IGF1 and *EGP* (environmental genome project) for IGFBP3. European population, allele frequency cut-off of *5 %* and tagSNP data coverage of *95%* were selected. All other parameters were left at default values, including r^2 threshold of *0.8*, LD range of *1.0-0.1*. *Genotype displays* were used to access local SNP positions, sample identifiers and corresponding genotyping allele designations, while *SNP summaries* allowed access to SNP reference sequence IDs, minor allele frequency, heterozygosity, chi-square statistics, function and conservation scores for each SNP. This data was saved as text files and converted to Haploview compatible formats .ped format and .txt files using locally designed software [106].

Genomic data for IGFBP1 was also accessed using the GVS *EGP* panel as before, however the European population selection was unavailable and *All* was stipulated for this parameter. This returned PDR90 (Polymorphism discovery resource 90) mixed panel genotypes, derived from 24 European Americans, 24 African Americans, 24 Asian Americans, 12 Mexican Americans and six Native American individuals. Although individual ethnic identifiers are unavailable, a study by Al Zahrani *et al.,* exploited the known increased genotype and haplotype diversity in individuals of African American descent in order to deselect these samples [50]. These same 28 samples (P002, P003, P004, P011, P015, P017, P018, P021, P032, P038, P039, P041, P042, P047, P050, P057, P059, P061, P064, P065, P074, P075, P076, P080, P082, P087, P088, P089) were manually removed from IGFBP1 data tables prior to Haplotype and tagSNP analysis.

The identity of a GVS "unknown" SNP (chromosome position 45900628, allele G/C) was identified as rs9658231 using dbSNP and the corresponding text altered prior to TagSNP selection to allow inclusion. Files were converted to Haploview compatible .ped and .info files as before.

2.2.2 Haplotype Construction and tagSNP Identification

Haploview was downloaded from http://www.broad.mit.edu/mpg/haploview/download and installed locally. IGF1, IGFBP1 and IGFBP3 genotype data files (.ped and .txt) were uploaded and Haploview used to estimate haplotypes and linkage disequilibrium. HW p-value of *0.01* and minor allele frequency (MAF) values of *0.01* were stipulated, all other parameters were left at default values. Haploview's tagger was used to section SNPs meeting the specified criteria into tag and capture bins.

2.2.3 Functional Impact Prediction (PupaSuite)

All potential Tagger tag and capture SNPs were analysed using PupaSuite (http://bioinfo.cipf.es/pupasuite/www/) to identify those with putative deleterious functional impact designations. dbSNP reference sequence identifiers for all tag and capture SNPs were uploaded and the following parameters selected; only predicted pathological non-syn mutations, pathological mut (PMUT), prot. structure and dynamics (SNPeffect), cellular processing (SNPeffect), pathological mutations predicted by selective constraints (dN/dS) and *Mus Musculus* conserved regions. All other criteria were left at default values.

2.3 Results

2.3.1 Haplotype Construction and TagSNP Identification

The following section displays linkage disequilibrium plots (Figure 8, Figure 9 and Figure 10), haplotype displays (Figure 11, Figure 12 and Figure 13) and tag / capture SNPs identified (Table 4, Table 5 and Table 6) across candidate regions IGF1, IGFBP1 and IGFBP3.

Linkage Disequilibrium plot

Pairwise estimates of D' were generated and graphically displayed using Haploview. The standard D′/LOD colour scheme is used; with red indicative of strongest linkage and linkage disequilibrium (D'> 0.8, LOD>2) and white indicative of weak evidence for both. The values in each square indicates the pairwise D' value, when no figure is listed the value of D' is actually one. Haplotype blocks containing multiple SNPs in strong linkage and linkage disequilibrium are highlighted by a border, and block size (Kb) displayed. Relative distance between markers along with reference sequence SNP identifiers are displayed above the plot. Haploview LD plot colour key is described in Table 3.

Table 3: Haploview LD plot standard D′/LOD colour scheme.

	D' < 1	D' = 1
LOD < 2	white	blue
LOD ≥ 2	shades of pink/red	bright red

Haploview linkage disequilibrium Plots are displayed for IGF1, IGFBP1 and IGFBP3 (Figure 8, Figure 9 and Figure 10 respectively). IGF1 contains strongest linkage disequilibrium and two large haplotype blocks spanning 13Kb and 35Kb containing 11 and 18 SNPs respectively. Regions of strong linkage disequilibrium are displayed in pink and red. IGFBP1 displays lower linkage disequilibrium with two haplotype blocks spanning 1Kb and <1Kb, containing five and two SNPs respectively, while IGFBP3 displays the lowest level of linkage disequilibrium with just one haplotype block spanning <1Kb and containing three SNPs. The reduction in linkage disequilibrium and linkage is clearly evidenced by comparison of IGF1 plots and IGFBP3 plots which display predominantly red/pink and blue/white profiles respectively.

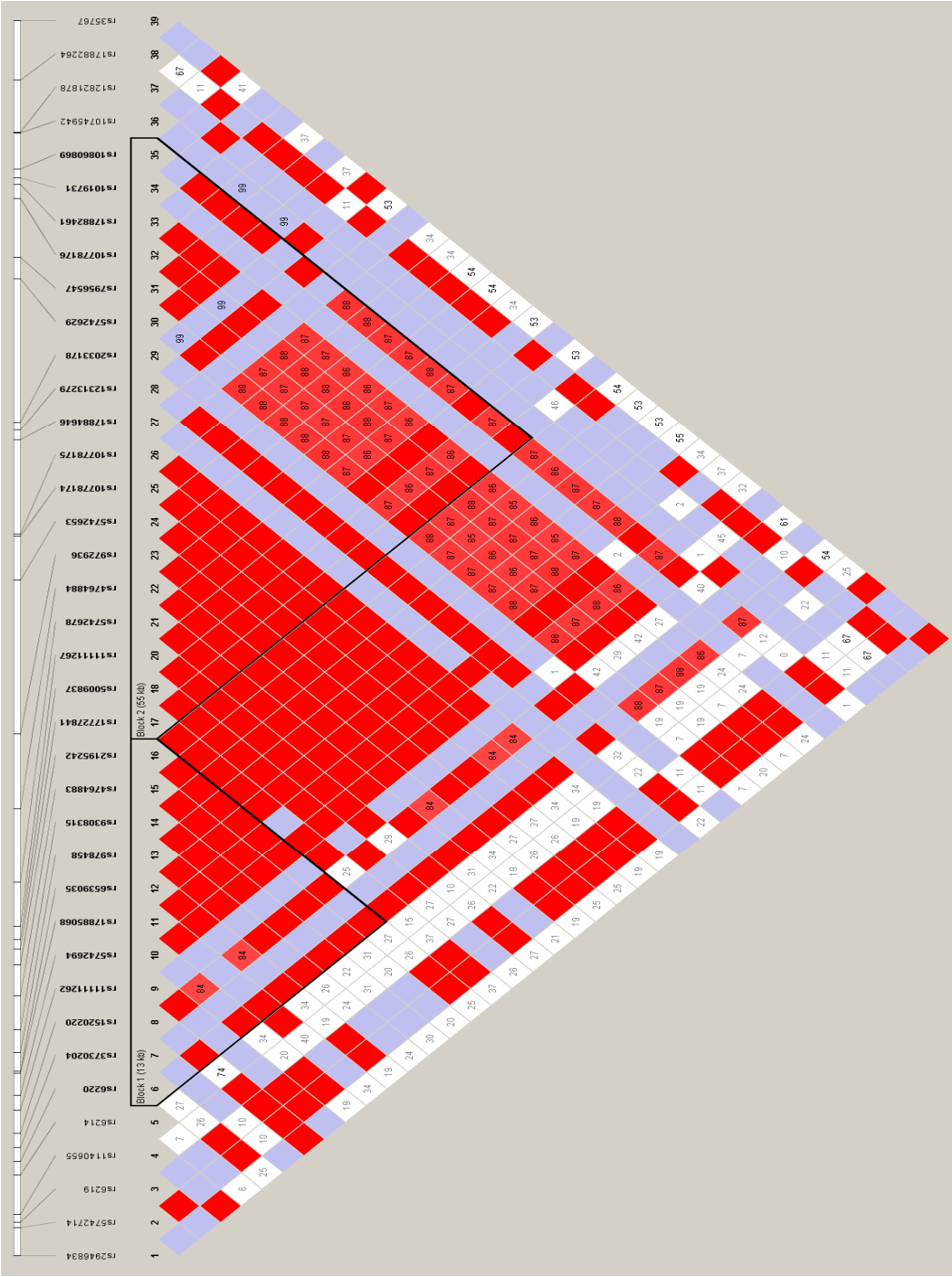


Figure 8 Haploview Linkage Disequilibrium Plot for IGF1. Two large Haplotype blocks spanning 13 and 35Kb, containing 11 and 18 SNPs respectively are visible. Regions of strong linkage disequilibrium are displayed in pink and red.

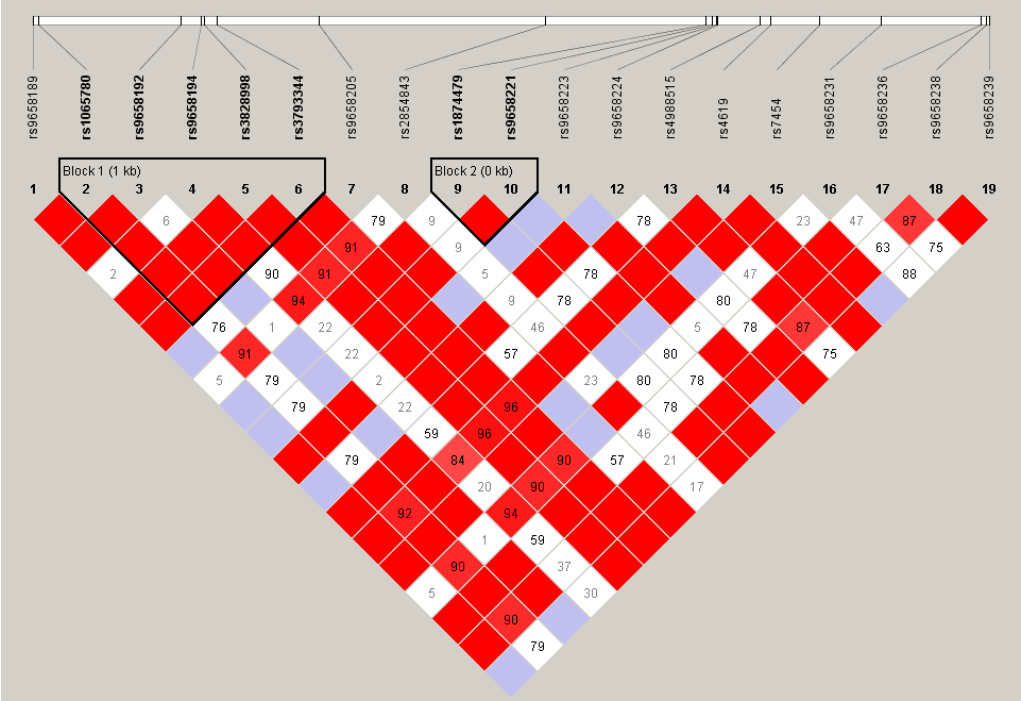


Figure 9: Haploview Linkage Disequilibrium Plot for IGFBP1 gene. Two haplotype blocks spanning 1 and <1Kb, containing 5 and 2 SNPs respectively are visible. Regions of strong linkage disequilibrium are displayed in pink and red.

Figure 10: Haploview Linkage Disequilibrium Plot for IGFBP3 gene. One haplotype block spanning <1Kb, containing 3 SNPs are visible. Regions of strong linkage disequilibrium are displayed in pink and red.

Haplotype Display

Multi-allelic D' was used to model linkage disequilibrium across candidate genes using Haploview and results displayed graphically. The horizontal sequential number list spanning block displays relates to the allocated SNP numbers (corresponding NCBI reference sequence identifiers are displayed on LD plots). Predicted haplotypes for each block are listed vertically from the highest to lowest population frequency and transitions between haplotypes in each block represented by lines whose thickness correspond to the population frequency. Hedrick's multi-allelic D', which represents the degree of LD (or recombination) between blocks, is displayed beneath transition lines [105]. Haploview derived haplotype displays for candidate genes IGF1, IGFBP1 and IGFBP3 are displayed in Figure 11, Figure 12 and Figure 13 respectively. Seven haplotypes are predicted for both IGF1 blocks, ranging from population frequencies of 0.326 to 0.022. IGFBP1 blocks are predicted to contain four and two haplotypes ranging from population frequencies of 0.653 to 0.065, while four potential haplotypes are predicted for IGFBP3 haplotype block, ranging from population frequencies of 0.386 to 0.023.

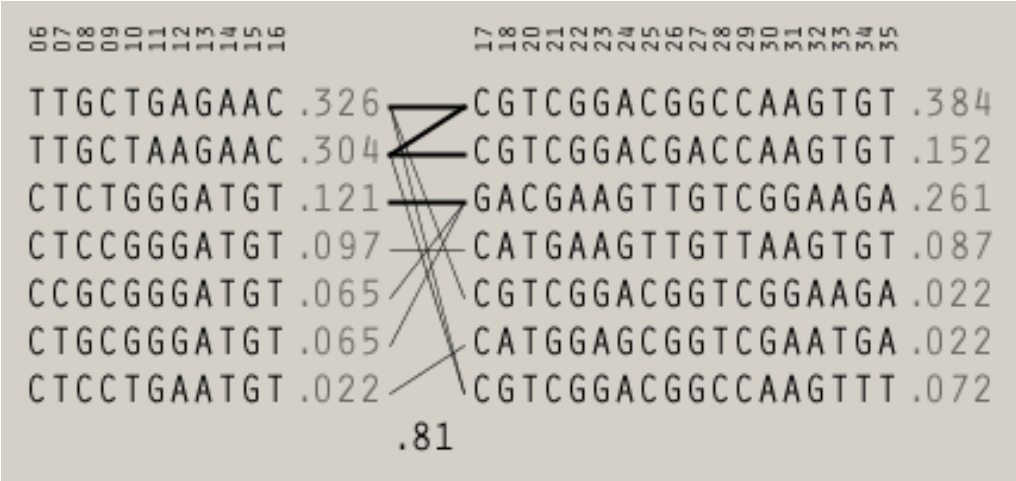


Figure 11: Haploview IGF1 Haplotype Display. Seven haplotypes are predicted for both IGF1 blocks, ranging from a population frequency of 0.326 – 0.022. D' = 0.81.

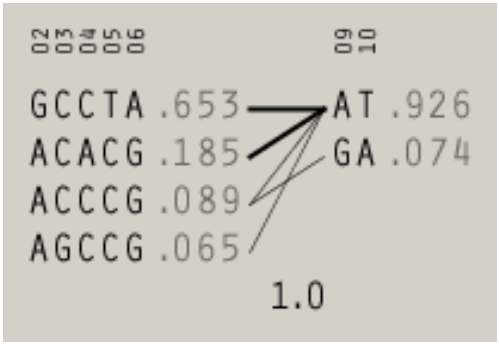


Figure 12: Haploview IGFBP1 Haplotype Display. Four and two haplotypes are predicted for 1Kb and <1Kb haplotype blocks respectively, ranging from a population frequency of 0.653 – 0.065. D' = 1.0.

Figure 13: Haploview IGFBP3 Haplotype display. Four potential haplotypes are predicted for this block, ranging from a population frequency of 0.386 – 0.023.

Tag and Capture SNP tables

The following tables; Table 4, Table 5 and Table 6 display all tag and capture SNPs identified by Haploview's implementation of Tagger for IGF1, IGFBP1 and IGFBP3 respectively.

Haploview's Tagger sections SNPs into bins of high linkage disequilibrium. Any single tagSNP in each bin may be used to capture all tag and capture SNPs with user defined correlation coefficients within that bin. Capture TagSNPs do not display high correlation with all bin-constituent SNPs and should therefore not be selected for analysis. IGF1 contains 35 potential tagSNPs and four capture SNPs sectioned into 13 tagSNP bins. Bin one contains the largest pool of 16 tagSNPs from which to choose, while bins seven to 13 inclusive contain self-tagging SNPs. IGFBP1 contains 19 alleles which were sectioned into 5 bins containing 14 tagSNPs and five capture alleles. IGFBP3 contains 25 SNPs sectioned into 14 bins; nine of these bins are self-tagging.

Testing a single tagSNP from each of the bins should allow for 100% variation capture across target regions with mean r^2 of 0.973, 0.957 and 0.968 for IGF1, IGFBP1 and IGFBP3 respectively. All tagSNPs listed capture all respective tag and capture SNPs with an $r^2>0.8$.

Table 4 IGF1 - Tag and Capture SNPs. Haploview's Tagger sections SNPs into bins of high linkage disequilibrium. Any single tagSNP in each bin may be used to capture all tag and capture SNPs with user defined correlation coefficients within that bin. Capture TagSNPs do not display high correlation with all bin-constituent SNPs and should therefore not be selected for analysis.

Bin No	TagSNPs	Additional Capture SNPs
1	rs12313279, rs10778175, rs10778174, rs5742653, rs972936, rs4764884, rs5742678, rs17885477rs5009837, rs2195242, rs4764883, rs9308315, rs978458, rs6539035, rs5742694, rs6220	
2	rs17882461, rs7956547 rs11111267	rs10860869, rs10778176, rs5742629, rs17727841
3	rs17882264, rs11111262, rs6219 rs5742714	
4	rs10745942, rs2033178	
5	rs1140655, rs2946834	
6	rs35767	
7	rs12821878	
8	rs1019731	
9	rs17884646	
10	*rs12316064	
11	rs1520220	
12	rs3730204	
13	rs6214	

* rs12316064 was formally listed as rs17885068

Table 5 IGFBP1 - Tag and Capture SNPs. Haploview's Tagger sections SNPs into bins of high linkage disequilibrium. Any single tagSNP in each bin may be used to capture all tag and capture SNPs with user defined correlation coefficients within that bin. Capture TagSNPs do not display high correlation with all bin-constituent SNPs and should therefore not be selected for analysis.

Bin No	TagSNPs	Additional Capture SNPs
1	rs7454	rs9658189,rs9658223, rs4988515, rs9658192, rs9658236
2	rs1874479 rs9658239 rs9658221 rs9658205 rs9658224	
3	rs3828998 rs3793344 rs1065780 rs4619	
4	rs9658194 rs2854843 rs9658231	
5	rs9658238	

Table 6 IGFBP3 - Tag and Capture SNPs. Haploview's Tagger sections SNPs into bins of high linkage disequilibrium. Any single tagSNP in each bin may be used to capture all tag and capture SNPs with user defined correlation coefficients within that bin. Capture TagSNPs do not display high correlation with all bin-constituent SNPs and should therefore not be selected for analysis.

Bin No	TagSNPs	Additional Capture SNPs
1	rs3793345	rs34678704, rs2471551 rs35440925, rs34091405
2	rs33979592	rs2854747, rs6413441 rs3110697
3	rs2132572 rs10255707	rs2132570
4	rs2132571 rs13241830	
5	rs2854744 rs2854746	
6	rs11765572	
7	rs12540724	
8	rs2453840	
9	rs2453839	
10	rs35751739	
11	rs6670	
12	rs13223993	
13	rs10282088	
14	rs34087654	

2.3.1 PupaSuite *in silico* Functional Prediction

The following section describes *in silico* functional analysis results obtained using PupaSuite. PupasSuite scores and brief descriptions are shown where appropriate. Deleterious predictions were not made for MUT, protein structure and dynamics, cellular processing, codon conservation, TFBS and I/E boundary analysis for any tag or capture SNP analysed and therefore output is not displayed. A number of deleterious predictions were made for IGF1 and IGFBP3 exonic splice enhancer SNPs (Table 7). Three SNPs (rs6214, rs6219 and rs2854746) were predicted to impact a deleterious effect with scores of -2, -2 and -1 respectively and although located in a srp55 responsive site IGFBP1 SNP rs4988515 was predicted to be neutral. IGF SNP rs5742629 was determined to be located within a triplex forming oligonucleotide target sequence, however predictions regarding the functional impact of such SNPs is not facilitated by PupaSuite at this time (Table 8).

Table 7: PupaSuite Exonic Splice Enhancer (ESE) Prediction for candidate genes IGF1, IGFBP1 and IGFBP3. SNP and transcript identifiers, SR protein types and prediction and associated scores are listed.

	IGF1		IGFBP1	IGFBP3
SNP ID	rs6214	rs6219	rs4988515	rs2854746
Transcript ID	IGF1-001	IGF1-001	IGFBP1-001	IGFBP3-201
SR Protein type	srp40	srp40	srp55	sf2
Allele scores	-2.73 (G) - 0.35 (A)	- 4.35 (G) - 1.97 (A)	-6.14 (C) - 4.59 (T)	- 3.56 (C) - 1.65 (G)
Prediction	Lose (-2)	Lose (-2)	Maintain	Lose (-1)

Table 8: PupasSuite Triplex Forming Oligonucleotide Target Sequence (TTS) Identification. TTS constituent SNP identifiers, genomic environment and associated TSS sequence are listed.

	IGF1	IGFBP1	IGFBP3
SNP ID	rs5742629	☒	☒
Genomic environment	Intron		
TTS Sequence	AAAGGAAAAAG		

2.4 Discussion

This chapter aimed to construct haplotypes across candidate genes IGF1, IGFBP1 and IGFBP3 and identify the minimal number of tagSNPs required to offer maximum variation coverage for Caucasian populations. *In silico* functional impact prediction was also performed to identify those SNPs most likely to impart a deleterious effect.

2.4.1 Haplotype Construction and tagSNP Identification

Using PGA (program for genomic application) genotyping data, 155 SNPs were identified in candidate gene IGF1. Of these 39 (31 intronic, four 3'UTR, two mRNA UTR and two intergenic SNPs) displayed a minor allele frequency of ≥5%. These were sectioned into 13 bins containing 39 potential tagSNPs using Haploview. Extensive LD was observed across the locus with two well-defined blocks spanning 35kb and 13kb. One SNP (listed as "unknown") chromosome base position 101363244 was identified as rs10778175 through dbSNP. Another listed as rs17885068 in Haploview has now been merged into rs12316064 and is listed using this identifier. Five SNPs (rs5742678,

rs5742694, rs1520220, rs6220 and rs2946834) previously found to correlate with altered IGF-I levels (Al-Zahrani et al., 2006) were included as either tag or capture SNPs. Using parameters as previously described; 39 of the 39 alleles (or 100 % of variation) may theoretically be captured with a mean r^2 value of 0.973 using 13 tagSNPs in 13 tests. A total of 100% of captured alleles show an r^2 >0.8.

IGFBP1 genotype data for individuals of European descent only was unavailable from NIEHS, however the known increased genotype and haplotype diversity in individuals of African American descent was utilised in order to deselect individuals of African descent from the PDR90 (Polymorphism discovery resource 90) mixed panel prior to Haplotype analysis. From "treated" PGA data, 63 SNPs were identified using Haploview. Nineteen had a minor allele frequency of ≥5% in the 64 PDR90 subjects of putative non-African descent (ten intronic SNPs, one mRNA UTR SNP, one coding-synonymous, one coding non-synonymous, one near 3' SNPs, three intergenic SNPs and two near gene 5' SNPs). LD was observed across two well-defined blocks collectively spanning >1kb. One non-synonymous SNP rs4619 previously associated with diabetic nephropathy was captured by the selection [107].

Using previously specified parameters, Haploview in combination with Tagger selected 5 tagSNPs, capturing 19 alleles (100% of variation) with a mean r^2 value of 0.957, using 5 tagSNPs in 5 tests.

Using IGFBP3 EGP (Environmental Genome project) genotyping data 74 SNPs were identified. Of these, 25 polymorphisms had minor allele frequencies of ≥5% (seven

intronic SNPs, two near gene 3' SNPs, one coding-synonymous, one coding non-synonymous, four intergenic, three mRNA UTR and one 3'UTR SNP). The locus exhibits low LD with just one haplotype block of <1kb encompassing three SNPs identified. Two SNPs (rs2132571 and rs2132572) previously found to correlate with alternate IGFBP3 levels were included as tag and capture SNPs respectively (Al-Zahrani et al., 2006). Using parameters as specified 25 of 25 SNPs (100% of variation) could be captured with a mean r^2 of 0.968 using 14 SNPs in 14 tests.

2.4.2 Functional Impact Prediction (PupaSuite)

All tag and capture SNPs were investigated by PupaSuite in order to identify those located in transcription factor binding sites, intron/exon border consensus sequences, exonic splicing enhancers and exonic regions – with putative deleterious effects. Although selection was not made on this basis, downstream experimental incompatibility may require incomplete bin representation. In instances where incompatibility is noted, the results of putative functional impact predictions will be used for prioritisation of bins. SNPs located in triplex-forming oligonucleotide target sequences (TTS) were also identified along with their location, however predictions regarding TTS impact could not be made at this time [108,109].

Disruptions to pre-mRNA processing mechanisms are of great interest to genetic disease research with up to 50% of all point mutations responsible for genetic disease inducing aberrant splicing [110]. Disruption of exonic splice enhancers (ESEs) are known to effect exon recognition, resulting in exon skipping altered expression and production of other

malformed splicing combinations [111]. Deleterious PupaSuite ESEs are identified using threshold scores derived by Cartegni *et al.,* SNP alleles causing significant deviation above set thresholds are more likely to disrupt normal ESE function causing aberrant effects. Three SNPs (IGF1 rs6214 and rs6219 and IGFBP3 SNP rs2854746) received deleterious predictions using this method. IGF1 candidate tagSNPs; rs6214 and rs6219 were found to be located in SRp40 responsive ESEs. The score difference threshold for this ESE type is >2.670; both received scores of 2.38 and designations of "lost" indicating a potential propensity toward mRNA processing misfunction [112].

The exact role of SRp40 is not yet fully described however these regions have been linked to development of type 2 diabetes and a reduction in insulin-dependent glucose uptake by insulin induced alternate splicing in Akt2 deficient mice. Here SRp40 was shown to reduce protein kinase C (PKC) ßII isoform expression via the PI-3 kinase signalling pathway [113,114]. Reduced PKC ßII expression, induced by selective mutation of SRp40 Ser86 residue (Serine > alanine) further demonstrated the significance a single amino acid change may have to splice site selection [113].

IGFBP3 capture SNP rs2854746 was found to be located within an SR Sf2 responsive ESE. ESE-Sf2 sites have a significant threshold score of 1.956. Rs2854746 SNP alleles generated a score difference of 1.91 and a designation of "lost". Sf2s act to maintain splicing regulation and accuracy and also to prevent exon skipping. Sf2- / ESE-mediated alternate splicing has yet to be established for IGFBP3, however a study by Smith et al., found that increased ESE-mediated Sf2 expression promotes splicing of IGF1 alternative

exon 5 in vivo [115]. Disruption of Sf2 ESEs have not as yet been linked with NIDDM however Sf2-specific ESE disruption has been linked with other disorders including spinal muscular atrophy [116].

IGFBP1 capture SNP rs4988515 (tagged by rs7454) was found to be located in an SRp55 responsive ESE. These sites have a significant threshold score of 2.676. A less severe score difference of 1.55 between alleles was seen in this instance and as such a verdict of maintain was calculated; indicating that the SNP is less likely to have a consequential effect on ESE recognition and functionality. It may also be of interest to note that Ensembl transcripts predict just one isoform for IGFBP1, therefore the ESE may be erroneously identified by PupasSuite or may work in conjunction with other sequences not currently recognised. One capture SNP rs5742629 (tagged by rs7956547) was found to be located within a triplex-forming oligonucleotide target sequence (TTS). Regulatory TTSs are concentrated heavily in promoter regions and play a role in expression regulation. By binding complementary triplex-forming oligonucleotides (TFOs), TTS RNAs form triple-helical oligonucleotide structures which have been shown to inhibit expression in multiple genes, including that of IGF-IR [117]. No method for assessing the functional impact of these SNPs has been developed yet however, and as such, a prediction regarding the likely impact of either SNP could not be made [118,119].

IGF binding protein post-translational modifications (including glycosylation, phosphorylation and proteolytic fragmentation) are known to heavily modulate IGF-I binding and bioavailability, as such it was hoped that PupaSuite's use of SNPeffect for identification of aberrant cellular processing of this type may prove particularly useful

with regard to analysis of candidate genes IGFBP1 and IGFBP3 [120]. No putative deleterious predictions of this SNP type were made however. Non-synonymous coding sequence SNPs were heavily prioritised in the past regardless of functional study beyond an awareness of an amino acid residue change. The techniques used for protein prediction have improved greatly and PupaSuite offers an excellent range of programs based on protein structure and dynamics, cellular processing and codon conservation prediction. A total of 12 coding non-synonymous and 6 synonymous SNPs have been identified across candidate genes IGF1, IGFBP1 and IGFBP3. Just four of these (two non-synonymous SNPs; rs4619 and rs2854746 and two synonymous rs4988515 and rs2132572) proved to have a minor allele frequency of 0.05 or greater in the target population.

Non-synonymous missense SNP rs2854746, was previously found by Patel et al., to confer a 12% change in circulating IGFBP3, and codes for proline or alanine depending on the nucleotide present (cytosine and guanine respectively) [55], while IGFBP1 non-synonymous rs4619 contains either alanine- or guanine-containing alleles. Some confusion abounds regarding the reading frame of the IGFBP1 exon; therefore the amino acid substitution created for rs4619 is unclear, methionine / isoleucine or tyrosine / cysteine may be produced by nucleic acid variants alanine and guanine respectively, depending on which reading frame is used. Missense mutations account for 48% of all reported human disease-causing alleles, however missense translations do not necessarily have any appreciable effect on protein function if a similar amino acid is substituted in place of the wild type residue [121]. Studies to assess the protein structure, binding and catalytic impact of such substitutions may be carried out to determine whether an

aberrant effect is likely to be imparted. This type of analysis is most effective if 3D structures are available, however structures for IGFBP3 or IGFBP1 proteins are currently unresolved (Protein Database, 07-08-07).

PupasSuite analysis found neither synonymous SNPs (rs4988515 and rs2132572) to incur a deleterious effect: IGF1 rs4988515 appears in the least highly conserved third-base codon position (TGC /TGT), however due to degeneracy the appearance of either C / T allele does not cause an amino acid substitution and cysteine addition is retained. Some confusion regarding the designation of SNP rs2132572 exists however, with GVS listing the SNP as intergenic while dbSNP build 36.3 (http://www.ncbi.nlm.nih.gov/projects/SNP/) predicts its location to be within an exonic region. According to the dbSNP designation, SNP rs2132572 is located in the third base position of serine codons TCC / TCT. A 30 kb haplotype block containing SNP rs2132572 was found previously by Al-Zahrani *et al.,* to affect plasma IGFBP3 levels [50]. An interesting question regarding codon bias is raised by synonymous polymorphisms. Codon bias relates to the unequal use of synonymous codons in an organism or gene family [122], a selective driver of which is thought to be translation efficiency. Cysteine TGT and serine TCT codons are known to facilitate G/U wobble, a feature which is less preferable than codons facilitating Watson-Crick pairing (i.e.TGC). In this way synonymous codons may yet be found to play in disease predisposition [123,124]. Studies by Kotlar *et al*., did not find any significant bias for general cysteine codon use across 16,000 human genes, however no such studies with regard to the IGF gene family have been carried out to date [123].

This work in this chapter set out to capture the variation across candidate regions IGF1, IGFBP1 and IGFBP3 using linkage disequilibrium based tagSNP selection for target SNP reduction with minimal loss of information as described in our project objectives "*to construct haplotypes for a given population and perform tagSNP selection which captures maximum variation across candidate genes IGF1, IGFBP1 and IGFBP3*". We also assessed all tag and capture SNPs in terms of their putative functional impact. Both of these issues were addressed; haplotypes were constructed across candidate genes and all SNPs meeting relevant selection criteria sectioned into bin to allow downstream tagSNP selection. All tag and capture SNPs were also assessed using PupaSuite and three SNPs predicted to impart deleterious impact. Theoretically any tagSNP from a single bin may be used to represent variation within that region of high linkage disequilibrium, however multiplex analyses (PCR and hybridisation assays) are to be used for amplification and genotyping of these targets, therefore all potential tagSNPs must be analysed in terms of their suitability for singleplex PCR amplification and probe hybridisation as well as multiplex compatibility with other species across linkage disequilibrium bins. Where bin representation may not be facilitated due to multiplex incompatibility, selection of tagSNPs will be directed by putative functional impact results where appropriate. The following chapter *primer and probe design and in silico analysis* aims to assess tagSNPs in terms of their suitability for multiplex PCR amplification and suspension-array probe hybridisation using a host of *in silico* techniques.

Chapter 3: Chapter 3: Primer and probe design and *in silico* evaluation

3.1 Overview

Chapter 2 *Haplotype construction and tagSNP analysis* facilitated organisation of tagSNPs into bins of strong linkage disequilibrium for maximum variation capture across target genes IGF1, IGFBP1 and IGFBP3 with minimal theoretical tagSNP requirements. If the benefits of this reduced resource are to be truly felt however, the probability for successful downstream amplification and genotyping must be maximised. Singleplex PCR and certain SA techniques are reasonably well-established; however, analysis of target sequences within the proposed multiplex format requires design of novel PCR primers and SA probes, which contribute sequence-induced variability to the assay. Careful planning, parameter selection and *in silico* evaluation of primer and probe sequences during the experimental design phase can greatly improve the chances of a successful outcome. This chapter provides an introduction to the methods which underlie PCR amplification and SA genotyping; experimental considerations are introduced with particular emphasis on robust primer and probe design. The techniques employed during the course of this work for construction and evaluation of IGF primer and probe sequences are discussed, and the final primer and probe set profiles described and critiqued.

3.1.1 Polymerase Chain Reaction (PCR)

The polymerase chain reaction has revolutionised the world of modern molecular biology by allowing rapid target specific amplification of nucleic acid sequences which facilitates robust analysis and sequence characterisation with a precision and sensitivity previously inaccessible to the wider scientific community.

The technique is derived from the naturally occurring cellular reaction, which uses DNA polymerase enzymes for replication and repair of DNA and RNA sequences *in vitro*. This process is the cornerstone of cellular replication and expression, a fact, which is reflected in the highly conserved nature of polymerase genes across species. The PCR concept was initially investigated by Kleppe *et al*., in 1971 who demonstrated repair and replication of short synthetic DNA sequences pertaining to the major yeast alanine transfer RNA gene *in vivo* [125], however the significance and widespread applicability of this technique was not wholly realised until 1983 when Karry Mullis devised an *in vitro* derivative [126]. Vital to increased PCR utility was the discovery of thermostable polymerase enzyme; unlike previous polymerases derived from mesophilic bacteria and bacteriophages, *Taq* DNA polymerases are purified from a chemotrophic thermophilic bacterium *Thermus aquaticus* that thrives at high temperatures (50°C to 80°C). The derived polymerase enzyme can itself withstand repeated heating to 95°C without significant loss of activity. This negates the need for stepwise enzymatic additions following DNA denaturation and has paved the way for the development of automated thermal-cyclers and high-throughput PCR amplification [126].

The significance of the *in vitro* PCR conception, which Mullis attributes to an "improbable combination of coincidences, naïveté and lucky mistakes", was recognised in 1993 when he was awarded the Nobel Prize in Chemistry [126,127].

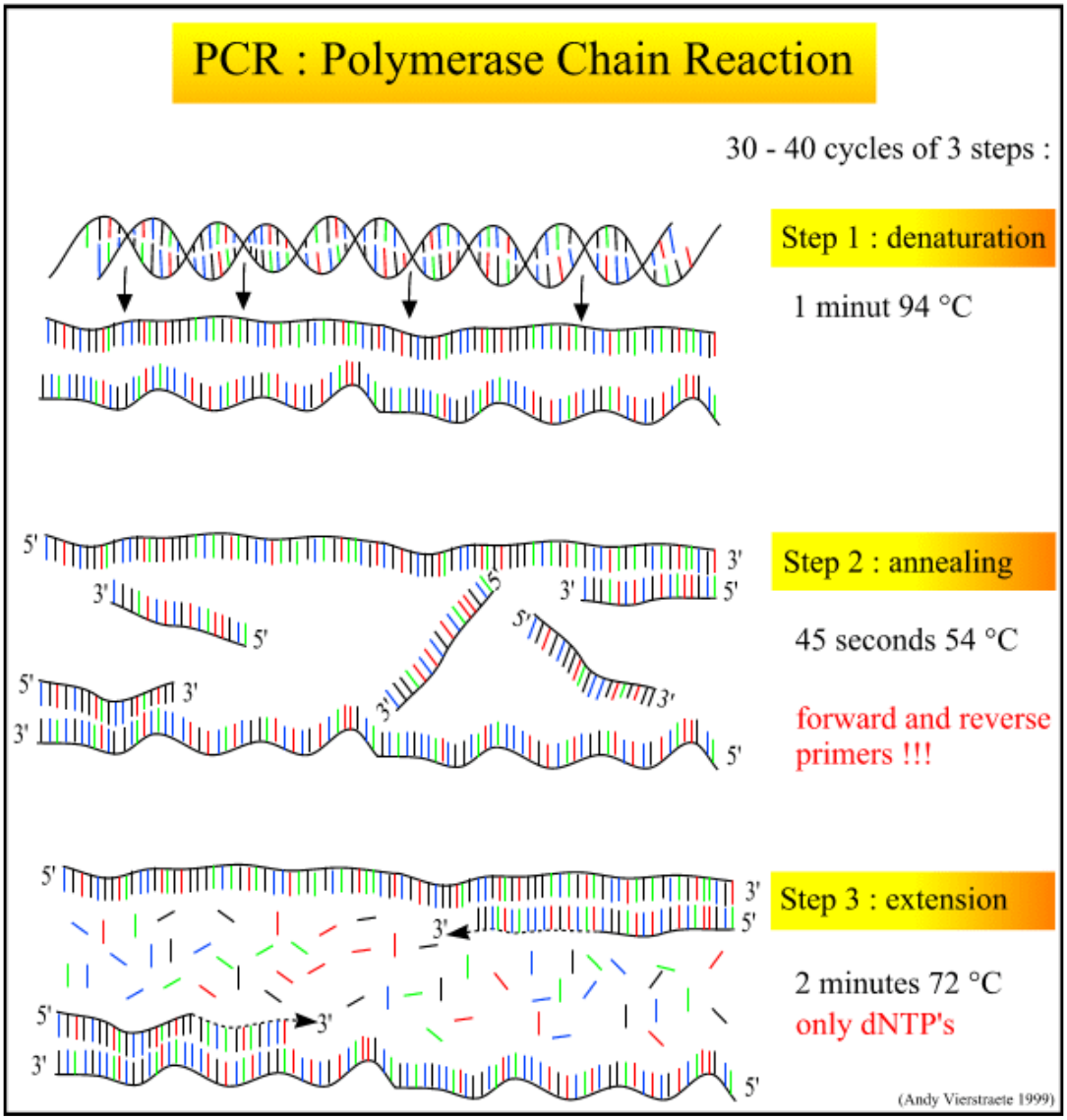


Figure 14: The three basic steps of the polymerase chain reaction are illustrated. High temperatures ~94°C are used to denature dsDNA into single strands, primers hybridise these targets during annealing ~54°C and increased temperatures ~72°C facilitate high polymerase activity and primer extension in the final extension step [128].

The standard application of this technique consists of three basic steps; denaturation, annealing and extension which are performed at ~94-96°C, 50-60°C and 68-72°C respectively.

- High temperatures are used to denature hydrogen bonds, which bind complementary double stranded DNA into single strands.

- Brownian motion causes the reaction components to move within the solution. Ionic bonds are formed between the synthetic single stranded oligonucleotide primers and nucleic acid template. Under suitably stringent conditions primers bind their complementary target sequences with increased stability allowing DNA polymerases to attach to ssDNA-primer duplexes.

- Once bound; DNA polymerases extend complementary sequences by sequential addition of free dNTPs to 3' primer ends in a manner directed by the complementary single stranded target.

Denaturation, annealing and extension steps are repeated between 25-50 times during standard PCR reactions [129,130].

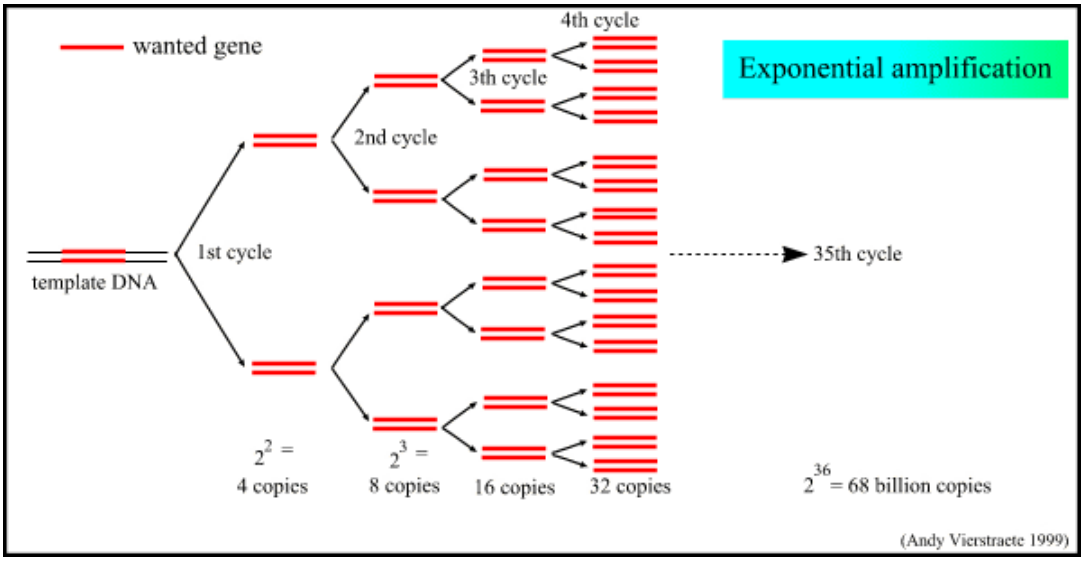


Figure 15: PCR amplification profile: The target gene is amplified from a double stranded target. The reaction is exponential with a single copy of each target produced with every PCR cycle assuming complete reaction efficiency. As such 4, 8, 16, 32 and 68 billion copies of this gene are produced following 1, 2, 3, 4 and 35 cycles respectively [128].

PCR amplification is an exponential process, with a single copy of each amplicon produced per cycle assuming complete reaction efficiency. Thus from a double stranded target; 4, 16, and 1,048,576 copies may theoretically be produced following one, two and 20 cycles respectively. In practice 100% efficiency is elusive however as limiting factors including diminishing dNTP concentration, temperature induced *Taq* inactivity and macromolecular crowding reduce final yields. Notwithstanding these limitations, large quantities of highly concentrated target sequence can be produced by PCR facilitating increased analytical sensitivity for evaluation of target sequences and other follow-on activities [126,129-131].

3.1.2 PCR Primer Design

Careful primer design is an essential first–step for high-efficiency, high-specificity PCR amplification. Specificity relates to the frequency with which mispriming or amplification of non-specific products occurs, while efficiency relates to the ability of the reaction to amplify products exponentially (i.e. doubling of targets per PCR cycle) and is reflected in yield. A large number of parameters including primer length, GC content, inter and intra-primer annealing temperature and propensity for cross homology must be considered if high functionality primers are to be developed [132].

Primer specificity is directed to a large extent by sequence length, GC content and applied annealing temperature. The genomic sequence of the target organism dictates the required minimum primer length, with increased genomic size inducing an increased primer length requirement. Allowing for some error, a minimum sequence length of 18

nucleotides is generally recommended for applications relating to the human genome, however for each additionally nucleotide used, primer sequences gain a four-fold increase in specificity. As such primer lengths of 18-24 nucleotides corresponding to annealing temperatures of >54°C are generally used [132]. Longer primers sequences of 28-35 bases have also been used to facilitate increased specificity for isoform-specific amplification and multiplex PCR amplification [133]. This design is implemented with some cost with regard to efficiency however, and amplifications of this nature require extended annealing times to facilitate hybridisation of primer-target duplexes [131,132].

During primer-target hybridisation lateral intra-strand hydrogen bonding occurs between complementary bases, however bonding strengths are not unilateral with adenine-thymine double bonds significantly weaker than guanine-cytosine triple bonds. As such, relatively balanced AT / GC nucleotide distributions should be applied to promote stable annealing along the primer length [134]. Also as polymerase effects elongation by addition of dNTPs to 3' ends, increased triple bond strength can be exploited by inclusion of G/C terminal nucleotides at ultimate and/or penultimate positions [135-138].

Both sequence length and GC content heavily influence primer-melting temperature (T_M). T_M is the temperature at which half of the DNA duplex will dissociate to become single stranded and is indicative of duplex stability. Annealing temperatures (T_A)[6] are derived from calculated T_M and must be high enough to guard against non-specific target hybridisation but low enough to facilitate target specific duplex formation. A number of other forces have also been found to impact duplex melting profiles including salt

[6] $T_A = Tm - 5°C$

concentration, the propensity for self-homology and base stacking forces (induced by the sequential order of nucleotides and vertical covalent bond formation along the length of the helical structure) [139,140]. The Panjokovich "consensus method" which calculates average predicted T_Ms from a number of comprehensive nearest neighbour T_M prediction algorithms and has been implemented in the primer design program Primer3 (http://frodo.wi.mit.edu/) and has been shown to display minimal error probabilities with regard to true T_M prediction [141]. It is essential that primer T_Ms be predicted accurately in order to minimise intra-primer pair T_M deviation. Divergent intra-primer pair T_Ms can result in use of suboptimal T_As, which prohibit efficient annealing of lower-T_M primers and induce non-specific amplification of the higher-T_M primers. This feature is especially important for multiplex reactions or standardisation of amplification within single labs [132].

If high-level primer functionality is to be maintained, primers should display homology with their intended target sequence only. Cross-homology facilitates identification of primers that display a propensity for strong homology within and between primers via hairpin, homodimer and heterodimer formation. These features can act to promote primer-dimer accumulation and inhibit target-specific amplification efficiency if left unchecked [142]. A stand alone program AutoDimer (http://www.cstl.nist.gov/div831/strbase/AutoDimerHomepage/DownloadPage.htm) has been developed by Vallone *et al*., to allow evaluation of primer sequences in this way [142,143].

A further primer-specificity check may be achieved by searching sequences against relevant target database (where such repositories are available). A number of programs including NCBI's Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and UCSC's PCR (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start) are specifically designed to search primer sequences against human genome assemblies and also provide relevant information regarding putative amplicon size and sequence. Detailed amplicon views also allow identification of sequence-based features including repeat regions and indels which may affect amplification specificity or amplicon size [144,145].

3.1.3 Suspension Array Genotyping

PCR acts to provide a concentrated yield of target molecules for improved sensitivity and specificity of downstream SNP genotyping assays. A wide array of divergent genotyping techniques have been conceptualised and investigated in an attempt to develop a flexible accurate, cost-effective, high-throughput system for sequence characterisation of this nature. These include enzyme based methods such as RFLP, primer extension and oligonucleotide ligase assays, post-amplification methods including SSCP gel and melting peak analysis and hybridisation discrimination approaches, including micro and suspension-arrays [146]. The impact of microarrays on SNP genotyping has been astonishing; using this method target specific capture molecules are spotted onto a planar array surfaces in a highly regulated manner which facilitates target identification by spot positioning. Hybridisation is subsequently carried out and fluorescent signals used to detect moieties captured under suitably stringent conditions. This approach facilitates characterisation of thousands of SNPs and even whole genomes in a single assay and has

been widely applied for disease SNP association and expression analysis. A number of limitations have, however, led to the development of suspension arrays that aim to alleviate some of these restrictive characteristics displayed by planar array formats [147,148].

Suspension arrays use suspended probe-coupled microspheres for target-specific capture and flow cytometric detection. Divergent optical microsphere properties (including particle size and fluorescence characteristics) can be used for identification of multiple probes within complex mixtures in a manner substitutive of physical location on planar arrays. This elevates problems associated with inert-spot, inter-chip variability and eliminates the need for complex spotting and dedicated data analysis instrumentation. Coupling of target molecules to their respective supports is also relatively straight-forward and can be performed in bulk to facilitate thousands of subsequent SAT assays [13]. Hybridisation of SAT target molecules mirrors that used for microarrays, however as SAT hybridisation is performed in suspension, binding kinetics are significantly improved and are more closely analogous to those observed in liquid-based hybridisation. The convex microsphere array surface also contributes to improved binding kinetics due to its reduced capacity for stearic hindrance [149]. SAT detection may be performed using a number of methods, although flow cytometry offers particle analysis rates of up to $10{,}000s^{-1}$, making serial multiplex analysis of this type extremely rapid [13]. Additionally a wide range of microsphere-encoding techniques have facilitated increased parallel analysis with multiplex detection of up to 100 codes detectable using the luminex^{100} flow analyser [150].

3.1.4 SAT Probe Design

Probe characteristics can greatly affect assay performance; probe type, length, GC content, intra- and inter-probe T_A, and the propensity for cross-homology all affect specificity, hybridisation efficiency and the capacity for probe allelic discrimination. As such, all contributory parameters must be carefully assessed during this experimental design phase if SAT allele discrimination is to be performed.

During SAT genotyping, allelic discrimination is achieved by hybridisation of PCR targets with allele-specific oligonucleotide probes. Under suitably stringent conditions, fully complementary sequences should display increased thermal stability relative to their non-complementary counterparts, thereby facilitating divergent hybridisation characteristics and increased hybridisation signals for fully complementary sequences [151]. Dissociation divergence of ~10°C may be seen for short oligonucleotide probes–target duplexes (11-17 bases) containing single base mismatches relative to fully complementary duplexes. This should be sufficient to allow discrimination between SNP alleles, however the extent of dissociation divergence is also influenced by a number of other factors. The position of the SNP within the probe sequence has been found by Letowski *et al*., to be important with respect to discrimination, with centrally positioned mismatches contributing greater instability than those located at non-central positions [152].

Selection of suitable hybridisation temperature (T_{HYB})[7] also plays an important role. Many published protocols still use "standard" 42°C hybridisation temperatures or hybridisation temperatures, which are significantly below predicted optimal T_{HYB} when performing suspension array genotyping. Letowski *et al.*, determined that altering hybridisation temperatures (8-13°C below predicted T_M) to meet the requirements of the probes results in increased specificity and increased true hybridisation signal determination [152]. As previously discussed, sequence derived factors which influence oligonucleotide melting and hybridisation temperatures include sequence length, GC content and nearest neighbour effects. Again nearest neighbour calculations provide the most accurate predictions and should be used for probe T_M/ T_{HYB} prediction [141].

Relative to PCR primer design, broader GC ranges of ~25-70% are permissible for hybridisation probes [10,153,154]. Use of highly divergent GC% in PCR based primers induce divergent annealing time requirements resulting in poor amplification efficiency or failure for highly divergent pairs. This range is relaxed somewhat for SAT probe design as allelic determination is made by relative inter-probe pair hybridisation signals. These probe pairs are predominantly composed of identical sequences and as such GC impact is likely to be of less consequence [10]. Probes with GC content below 20% may display poor hybridisation due to the reduced strength of AT double bonds relative to GC triple bonds, while those exceeding 80% GC have been shown to display drastically increased cross-hybridisation [154].

[7] T_{HYB} is the temperature at which annealing between probe and targets is performed.

The non-genomic PCR amplicon targets generally used for SAT mean that specificity can be achieved using smaller probe sequences although, whilst not strictly required, probe lengths of 15-25 bases are generally used and adjusted within this range to fit a universally applied hybridisation temperature [10,153].

As for PCR primer design, it is essential to evaluate all SAT probes fully in terms of their propensity for cross-homology. Due to the tethered nature of microarray probes, cross-homology analysis between alternate probes within this format is not required. Overlooking this feature within the scope of SAT probe design can result in coagulation of coupled microspheres and hybridisation failure. As such, evaluation of potential hairpin and homo/hetero-duplex structures using programs such as AutoDimer is recommended [10].

The following *in silico* primer probe design and evaluation methods were employed to address the thesis objective "*To perform multiplex primer and probe design, utilising in silico and manual analysis for evaluation and selection of a high specificity primer / probe sets in a manner compatible with downstream multiplex PCR and SAT analysis*".

3.2 Methodology

Using Primer3 (http://frodo.wi.mit.edu/), primer pairs and SAT probes for all tagSNPs generated by HaploView's Tagger (http://www.broad.mit.edu/mpg/haploview/download) were designed within strict parameter thresholds. Pairs were analysed for specificity using UCSC *(*University of California Santa Cruz*) in silico* PCR program and evaluated in terms of their constituent sequence type and insertion deletion polymorphisms. The proximity of probe-adjacent SNPs was determined for all probes, and proximity of primer pairs performed using Gene Infinities Primer Map (http://www.geneinfinity.org/sms_primermap.html).

Multiplex compatibility of primer pairs was assessed using AutoDimer (http://www.cstl.nist.gov/div831/strbase/AutoDimerHomepage/DownloadPage.htm) until a primer pair combination representing the largest number of bins across candidate genes IGF1, IGFBP1 and IGFBP3 was identified. The following section describes the methods and thresholds used to determine the final multiplex primer and probe sets [142,155].

3.2.1 Sequence retrieval

The Ensembl Homo sapiens search panel was accessed at http://www.ensembl.org/index.html and dbSNP reference sequence identifier used to retrieve corresponding FASTA format target SNPs and flanking sequences (position; + 200, -200 bases). In cases where downstream primer design could not be used to produce a suitable primer pair for amplification within this optimal target size range, flanking sequences of +400, -400bases were permitted.

3.2.2 Primer3 Primer and Probe Design

Primer3 was accessed at http://frodo.wi.mit.edu/ and used to design primers and probes for all tagSNPs. Both primers and probe sequences were designed concurrently to determine suitability of target to both (PCR and SAT) downstream applications.

Ensembl derived tagSNP sequences (with associated descriptor listing gene name, haplotype bin number and target SNP reference sequence identifier) were inputted and the following parameters specified for primer design: default minimum product size range 100-300bp, primer size: (min) 18 (opt) 29 (max) 35, primer Tm: (min) 64 (opt) 65 (max) 67, max Tm difference: 2.0, primer GC% (min) 40(opt) 50 (max) 60, Max poly X: 3 and CG Clamp: 1. Hybridisation probes were selected using the following specifications: Hyb Oligo Size: (min) 13 (opt) - (max) 40, Hyb Oligo Tm: (min) 55 (opt) 60 (max) 65 and Hyb Oligo GC%: (min) 25 (opt) - (max) 80. All other parameters were left at default values.

Primer3 returns primer pairs in order of PCR suitability given the specified parameters; as such the first primer pair returned in each search was initially selected and primer pair sequences and associated information retained for further *in silico* analysis. Where possible an attempt was made to choose primer pairs with amplicons of alternate sizes (≥7 base difference) to allow for subsequent separation and analysis of multiplex amplicons.

3.2.3 UCSC *in Silico* PCR

The UCSC PCR facility was accessed at http://genome.ucsc.edu/cgi-bin/hgPcr?org=Human&db=hg18&hgsid=91360718. Forward and reverse primer sequences designed using Primer3 were inputted and the human March 2006 assembly selected. All other parameters remained at default values. All primer pairs were assessed in terms of amplicon specificity and those producing more than one amplicon or no amplicon, replaced with alternate pairs where possible.

3.3.4 UCSC Amplicon Analysis

Detailed information regarding the nature of *in silico* predicted amplicons was accessed using UCSC Genome Browser (http://genome.ucsc.edu/). The following UCSC Genome Browser information was displayed: ref seq genes, humanESTs, human ESTs that have been spliced, SNPs (129), segmental dups, structural var, repeat masker, simple repeats, microsatellites, self chain. Each amplicon constituent non-target SNP was investigated in terms of the polymorphism type and proximity to primers and probe sequences. Amplicons containing known indels were replaced where possible. Where substitution could not be facilitated, indel amplicons were permissible. Polymorphisms within primer sequences and non-target polymorphisms within probe sequences were not permissible. Repeat region tagSNPs were also replaced with alternative tags where possible, however these SNPs were permissible where replacement could not be made.

3.2.5 Primer Pair Proximity Analysis

The chromosomal position of amplicons was plotted using the Gene Infinities' Primer Map program (accessed at http://www.geneinfinity.org/sms_primermap.html) to assess the position of primers and amplicons on the genomic strand. Full FASTA format gene sequences for IGF1, IGFBP1 and IGFBP3 were mined from Ensembl as before using gene names rather than reference sequence identifiers. Gene sequences and corresponding primer pairs were inputted and selections made as follows; maximum allowable bases per line;*105*, reading frame;*1* and restriction sites *not be shown*. Resultant Primer Map windows detailing the position of primers across the target sequence was evaluated and potentially problematic adjacent primer pairs identified. Screen capture was used to save results of interest; these include any adjacent primer pairs within 800 bases of each other.

3.3.6 AutoDimer Cross-homology Analysis

Primer pairs and probes were assessed in terms of their suitability for multiplexing by assessing their putative potential for dimer (homodimer and heterodimer) and hairpin formation using AutoDimer. AutoDimer was accessed at (http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm) and downloaded locally.

Files containing Primer3 primer pair sequences and identifiers were assembled and uploaded to the AutoDimer pane in FASTA format and *Temp for dG calc* of 64°C (putative optimal annealing temp -1°C) specified prior to analysis. Results were assessed

and dimers receiving threshold scores ≥7 replaced with alternate primer pairs representing the identical tagSNPs or bins. Primer pairs receiving scores of 7, which could not be substituted for less reactive alternatives, were assessed in terms of their predicted melting temperature and Gibbs folding free energy (delta G). Duplex's receiving predicted delta G >0 kcal/mol (at *64*) were deemed permissible. Those exceeding this threshold were removed regardless of whether bin representation was maintained or not. This process was repeated for probe sequences using a *Temp for dG calc* of *59* (putative hybrid temp -1°C).

3.3.7 Primer Pair and Multiplex Set Nomenclature

The final primer and probe sets were termed "IGF multiplex sets". For clarity primer pairs selected for inclusion in this final set were named using a single prefix and underscore-separated suffix. The prefix represents the tagSNP constituent gene, i.e. prefixes IGF1, BP1 and BP3 represent genes IGF1, IGFBP1 and IGFBP3 respectively, while the suffix represents the TagSNP bin number as conferred by Tagger. Probes were named using similar nomenclature using the gene name/bin number prefex/suffix as before, however an additional letter (A, T, C, G) or symbol (-) were also added to denote alleles adenine, thymine, cytosine, guanine and deletion respectively. TagSNP reference sequence identifier for each IGF multiplex set primer / probe are detailed in electronic appendix Table 1 and Table 6 respectively.

3.3 Results

3.3.1 Primer3 Primer Sequences

Primer3 was used to design primers for tagSNP targets spanning candidate genes IGF1, IGFBP1 and IGFBP3. Forward primers are designed to hybridise sense strands while reverse primers should hybridise their antisense counterparts. Primer sequences for IGFBP3, IGF1 and IGFBP1 are displayed in Table 9, Table 10 and Table 11 respectively. Haplotype bin numbers are displayed and 3' CG clamps highlighted in bold lettering. Reference sequence identifiers for tagSNP used in the final IGF multiplex set along with corresponding bin numbers and identifiers can be found in electronic appendix Table 1. It should be noted that all reverse primers contain a 5' biotin attachment to facilitate downstream streptavidin-fluorophore attachment.

Table 9:IGFBP3 primer sequences. Both forward and reverse primer sequences (5'→3') designed to hybridise with sense and antisense genomic sequences respectively are displayed. The Tagger derived bin numbers, which they represent, are also displayed.

Bin	Forward Primer 5'→3'	Reverse Primer 5'→3'
1	AACACGCTTATAAGAGCTTGGTGT**CC**	GAGTGGGACTTTGGCATTGTCT**G**
3	GACTCTGCTATGCTGAGAAAGCACAA**C**	GCTGGTGTCCACCTTATACTCCTAGAAA**C**
4	TACACCGCAAGTCTCCAATTAAGAGT**G**	CCAACGCATCGAGAATACAGTAATA**CG**
6	TGTCGTCTACAAGAACCAAGGTGT**G**	TCCACGAGGTACACACGAAT**GC**
7	CATCATACTACTCACTACATGGTGGTT**GC**	ATGAAGCATACTGCCTTCACCTACTGA**G**
9	GCCCTGAGATATCCAGCACA**GC**	TGAACACTGTGAGCAGCATCT**GG**
10/11	ATTACTTGTGATGCCTCTGAATGT**GG**	CTCTGGGAACCTATAAAGGCAGGTATTT**C**
12	GTGAGCTCCTTTCCTCAGTCAT**GG**	GAGATTCACCCATGTTTGTTGAACTTAGA**G**
13	ACACACCACAATACCAGTCCTCTGAA**C**	AGTCGAAGAGTTACCTCCTGTCGAGTTA**C**
14	GTTGATCATAGGTATTGTGTCAGGGTTT**C**	GATGGTGAGACTTAGCCTCCATACTTA**GC**

Table 10: IGF1 primer sequences. Both forward and reverse primer sequences (5'→3') designed to hybridise with sense and antisense genomic sequences respectively are displayed. The Tagger derived bin numbers, which they represent, are also displayed.

Bin	Forward Primer 5'→3'	Reverse Primer 5'→3'
1	GGGTCTCTTTCTCTTAGCCTTCTATCT**GG**	ATCTTTGGGTCTTGTTATAGTGCCTTCT**C**
2	ATACCTCAGCATTGGCAATAGATTCT**G**	TCTGCAACTTACTTGGTGAGTGATCTT**G**
3	GCTAAAGCACATTTGAGATTACACAGA**CC**	AGTCAGTACAAGATGTTGACCATCGACT**C**
4	GCTAAAGCTGGAATAATGTGTTAGGTGT**G**	GACAGTGATTTGCATGTAGAAAGTGCT**C**
5	CTCTATATCCCTGGGTGTTACCTGCATA**G**	GCACCTTTGAGTGATGACCTATTATTGA**G**
6	GCAGACATACCTCTTTCCCTAGAGA**GC**	TAACACAAAGAGCCAGAGTAGGATTTCAA**G**
7	AGTTGTCCAATATCCTTAAGTGTCTGT**GC**	ACTAGGTAATTGCCAAGCCTAGAAGTGT**C**
8	GCTTTCCCACAGCTAGTGACTGTA**CC**	TTAGACTGCCTGCTATGCATCTGT**G**
9	CTCTCACCTGCCACCATGTAAGATAT**G**	GTTTCACCGTGTTAGCCAGGAT**G**
10	CGTGTGCCTGTAGTTTCAGCTACTCA	GCCGTTGTTATTAGTCCTCAGTGATCTTT
11	GTTGAGCTAATAGAGAGCTTGAACCTT**GG**	TTACTAGGAAAGGATCTAGAGGCCAGAA**G**
12	CTGAAGTTCCTCTTGGAAGGCATAA**C**	AGATTCCATCTGTGGCATTTGTA**CC**
13	AGATAATATGGCAGTGCATCTTTCA**GC**	GAATAAGATACTGGACTCCTCTTCCCAA**G**

Table 11: IGFBP1 primer sequences. Both forward and reverse primer sequences (5'→3') designed to hybridise with sense and antisense genomic sequences respectively are displayed. The Tagger derived bin numbers, which they represent, are also displayed.

Bin	Forward Primer 5'→3'	Reverse Primer 5'→3'
1	ATTTCTGCTCTTCCAAAGCTCCT**G**	ATGGTGGAATATACAAGTTAACCGTCCT**C**
2	GAGTGCTTTAGGTCTCAGTGAAGTACA**GG**	GTGCAATAATGACTTCCCATGTGT**G**
3/4	CACAACTAGAGCTTGAAACCAGAGCA**C**	CTTCCTCCTTGAGTCTCCACTAAGCTAT**G**
5	TCCTGGAGACTCTAGCTCCCTATCTT**G**	TATCAGTCTGTCCCTGTCCCACAT**C**

3.3.2 Primer Parameter Profiles

Critical parameters which may affect efficiency and or specificity of PCR amplification were determined and profiled for each primer / primer pair. These parameters include amplicon length/ repeat regions, primer-annealing temperature and GC content (Figure 16, Figure 17 and Figure 18 respectively). Data is displayed in Matlab-derived plots for ease of interpretation and comparison between all pairs in the full IGF multiplex set.

Amplicon length of each primer pair was determined using UCSC in silico PCR. Predicted mean amplicon length was determined to be 209bp, spanning 428bp from 72 – 500bp (IGF1_8 and IGF1_12). Three targets (IGF1_9, IGF1_10 and BP3_14) contain repeat regions, spanning 100%, 30% and 87% of each amplicon respectively (Figure 16). The Panjokovich consensus method as implemented in Primer3 was used to predict T_A for all primers in the IGF Multiplex set. Mean inter-primer T_A is 65.24°C, spanning 1.34°C from 64.47 – 65.81°C (IGF1_12 and BP3_12 respectively), while mean intra-primer T_A is 0.548 °C, spanning a maximum of 1.32°C for primer pair BP3_10/11 (Figure 17). GC content for all primers in the final IGF set was determined. Mean inter-primer GC% was determined to be 46.31%, spanning 15% from 40.0–55.0% (IGF1_6R/BP3_12R and BP3_6R respectively), while mean intra-primer GC% is 4.33%, spanning a maximum of 14% for primer pair BP3_12 (Figure 18). Numerical table accompaniments for amplicon length/ repeat regions, primer-annealing temperature and GC content plots are detailed in electronic appendix Table 5, Table 3 and Table 4 respectively.

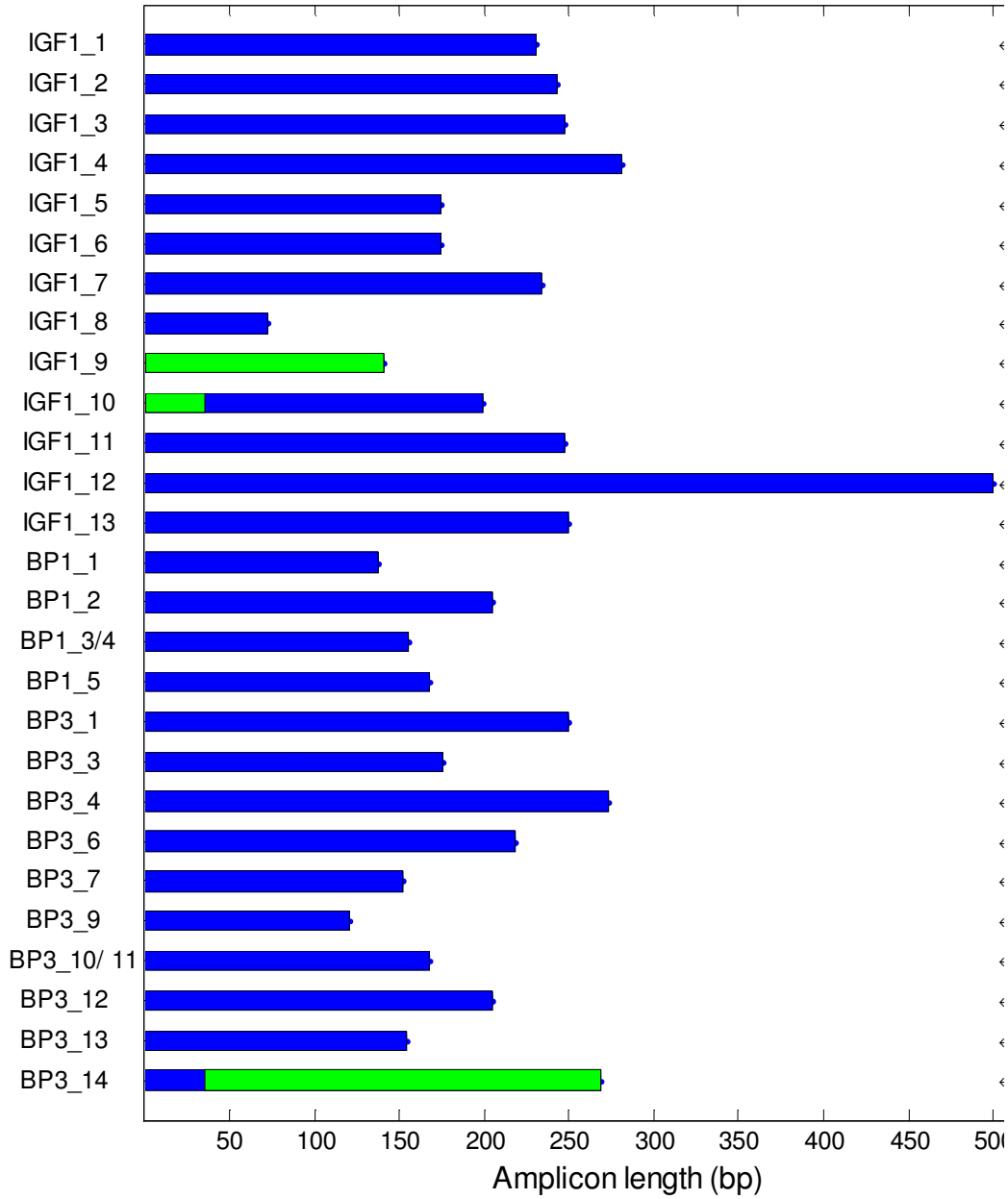


Figure 16: Amplicon profiles for the IGF multiplex set. Predicted amplicon length for each primer pair is displayed, with repeat and standard genomic regions highlighted in green and blue respectively. Predicted mean amplicon length is 209bp, spanning 428bp from 72 – 500bp (IGF1_8 and IGF1_12). Three targets (IGF1_9, IGF1_10 and BP3_14) contain repeat regions, spanning 100, 30 and 87% of each amplicon respectively.

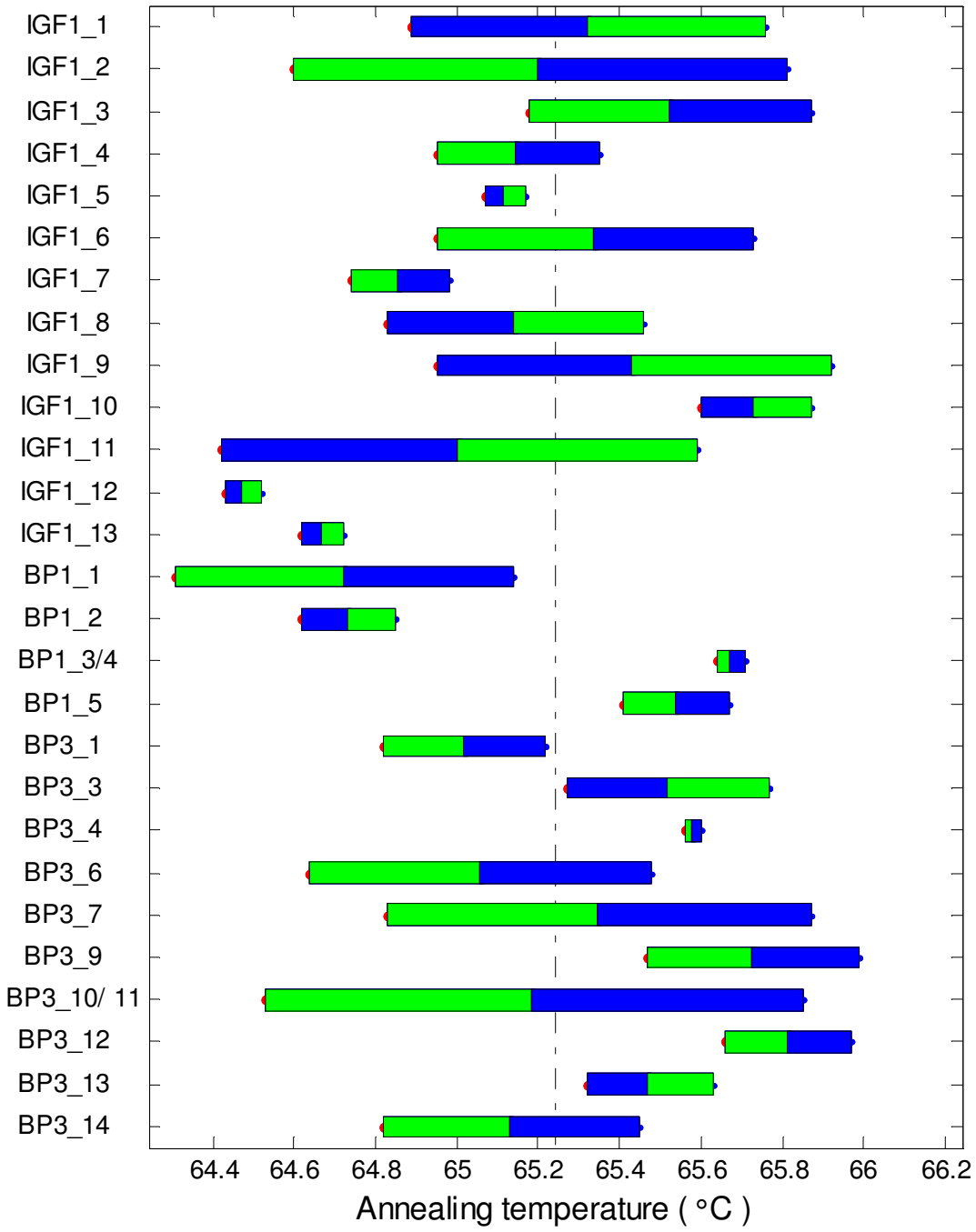


Figure 17: Primer3 (Panjokovich consensus) primer predicted T_A. The predicted T_A for each forward (blue) and reverse (green) primer is given for each pair. The vertical terminal box position denotes the predicted optimal T_A for each primer. The intra-primer vertical line separating forward and reverse primers in each pair denotes the predicted optimal T_A for that pair. The central dotted vertical line denotes that optimal predicted T_A for all primers in the full IGF set. Mean inter-primer T_A is 65.24°C, spanning 1.34°C from 64.47 – 65.81°C (IGF1_12 and BP3_12 respectively), while mean intra-primer T_A is 0.548 °C, spanning a maximum of 1.32°C for primer pair BP3_10/11.

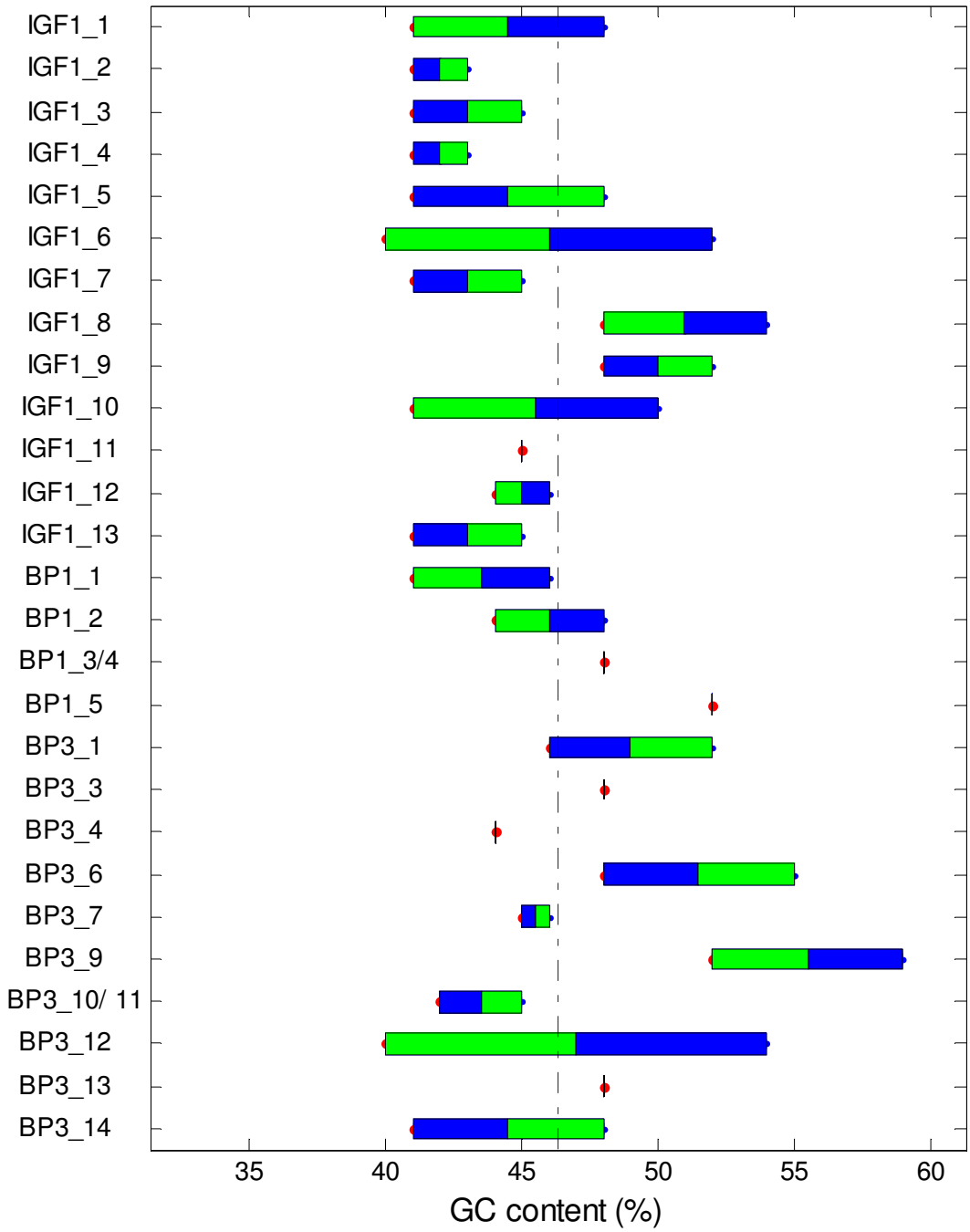


Figure 18: Primer GC content. GC content for all primers in the final IGF set are shown. GC content for each forward (blue) and reverse (green) primer are displayed. The vertical terminal box position denotes the GC % for each primer. The intra-primer vertical line separating forward and reverse primers in each pair denotes the mean GC% for that pair. The central dotted vertical line denotes the mean GC% for all primers in the full IGF set. Mean inter-primer GC% is 46.31%, spanning 15% from 40.0–55.0% (IGF1_6R/BP3_12R and BP3_6R respectively), while mean intra-primer GC% is 4.33%, spanning a maximum of 14% for primer pair BP3_12.

3.3.3 Primer3 Probe Sequences

Primer3 was used to design probes for tagSNP targets spanning candidate genes IGF1, IGFBP1 and IGFBP3. Probes were designed concurrently with primers using the *internal oligo probe* function and are designed to hybridise with antisense biotinylated PCR amplicon strands. Probe identifiers, SNP alleles and probes sequences for IGF1, IGFBP1 and IGFBP3 probe pairs are displayed in Table 12, Table 13 and Table 14 respectively. SNP alleles are highlighted in red. Further details regarding probe sequence identifiers and corresponding bin numbers may be found in electronic appendix Table 7.

Table 12: IGF1 probe identifiers, constitutive alleles and probe sequences are displayed. Probe alleles are highlighted in red.

Probe Identifier	TagSNP Allele	Probe Sequence
IGF1_1T	T	TTCTGCATTTCTCTGAATGTCAG
IGF1_1C	C	ACTTCTGCATTTCCCTGAATGT
IGF1_2A	A	AGATGAGAAAATTGAGGAACAAACA
IGF1_2T	T	GTTGAGAAAATTGAGGAACA AACA
IGF1_3G	G	CATAGGGATCGGCAGGTTT
IGF1_3A	A	CCATAGGGATCAGCAGGTTTT
IGF1_4A	A	TTCCCTATAGAGCTTGGCATTT
IGF1_4G	G	TTCCCTATAGGGCTTGGCAT
IGF1_5A	A	AATGACACATTATTAGATACATTGGTTACC
IGF1_5G	G	TGACACATTATTGGATACATTGGTTAC
IGF1_6A	A	TTTTCCACATGACTCTCAGGG
IGF1_6G	G	TTTCCGCATG ACTCTCAGG
IGF1_7G	G	AACATCATAG GCATAGAAAGATCCA
IGF1_7A	A	CAAACATCATAGACATAGAAAGATCCA
IGF1_8C	C	TAACTTTGACCAGCTGTCACACA
IGF1_8A	A	CTAAATTTGACCAGCTGTCACACA
IGF1_9C	C	GCGGATCACGAGGTTAGAAG
IGF1_9T	T	GGTGGATCAC GAGGTTAGAA GA
IGF1_10C	C	CCAGGAGGCGGAGGTT
IGF1_10T	T	AACCCAGGAGGTGGAGGTT
IGF1_11G	G	TTGCCAAACCTCACTCAGG
IGF1_11C	C	TTTCCCAAACCTCACTCAGG
IGF1_12A	A	GCACATTAACTCATCATTTGAAGG
IGF1_12G	G	CACATTAACTCGTCATTTGAAGGA
IGF1_13C	C	AAAACACGTTAAGTCTGCAGAAGA
IGF1_13T	T	CAGAAAACATGTTAAGTCTGCAGAAG

Table 13: IGFBP1 probe identifiers, constitutive alleles and probe sequences are displayed. Probe alleles are highlighted in red.

Probe Identifier	TagSNP Allele	Probe Sequence
BP1_1C	C	CGTCTGTTTT TAAAGAGCATGGA
BP1_1G	G	CGTCTGTTTT TAAAGAGGATGGA
BP1_2A	A	CTGCTTCACAGGCAATGAAC
BP1_2G	G	TGCTTCACGGGCAATGA
BP1_3T	T	CAGGACGTGCTCTGGGAG
BP1_3C	C	CAGGACGTGCCCTGG
BP1_4C	C	ATTGCACGGTCTTGGCAG
BP1_4A	A	CATTGAACGGTCTTGGCAG
BP1_5A	A	GCCAGGCTGCCATCC
BP1_5G	G	GGCTGCCGTCCTCTCTG

Table 14: IGFBP3 probe identifiers, constitutive alleles and probe sequences are displayed. Probe alleles are highlighted in red.

Bin no	TagSNP Allele	Probe Sequence
BP3_1T	T	GGCTCAGAATCATGCAAGC
BP3_1C	C	TCAGAATCACGCAAGCATGT
BP3_3T	T	AAGAGCCATGCGTGCCTA
BP3_3C	C	CCACGCGTGC CTAGG
BP3_4T	T	ATGGAGTTTACACCCATGACAAA
BP3_4C	C	GGAGTTTACA CCCACGACAA A
BP3_6G	G	AGCCGGTGTCGGGG
BP3_6A	A	GAGCCGGTGTCAGGGAA
BP3_7A	A	CTAAAGAAGGCAGACAAACGCT
BP3_7G	G	GAAGGCGGACAAACGCT
BP3_9T	T	GTCTCAACTCATGTTTTCAA ACAAA
BP3_9C	C	GGTCTCAACTCACGTTTTCAAAC
BP3_10C	C	GTCCCTCCTACCCCACG
BP3_10T	T	GTCCTTCCTACCCCACGG
BP3_11T	T	GACTCTCCCTGTCTCTCTGTCC
BP3_11-	-	GACTCTCCCGTCTCTCTGTCC
BP3_12T	T	CACAGTTGTATCATATAGCATCTCTAACAT
BP3_12A	A	ACAGTTGTATCAAATAGCATCTCTAACATT
BP3_13G	G	TTACAGAACCGGCTTGCTG
BP3_13A	A	TACAGAACCGACTTGCTGCTC
BP3_14C	C	CTATCATCTATCTAGTCTATCTACCTACTT ATCTC
BP3_14A	A	ATCTATCATCTATCTAGTCTATATACCTAC TTATCTC

3.3.4 Probe Parameter Profiles

Critical parameters including inter- and intra-probe pair melting temperature (T_M) and GC content which may affect efficiency and or specificity of probe hybridisation were determined and profiled for each probe / probe pair. For ease of interpretation all probes T_M and GC data is displayed in Matlab-derived plots (Figure 19 and Figure 20 respectively). Numerical table accompaniments are detailed in electronic appendix (Table 8 and Table 9 respectively).

The Panjokovich consensus T_M method as implemented in Primer3 was used to predict probe T_M. Mean inter-primer T_M for the full IGF Multiplex probe set was predicted to be 60.09°C, spanning a 3.49°C degree range from 57.96 to 61.45°C (BP1_3C and IGF1_8A respectively). The mean intra-primer T_M was predicted to be 0.548 °C, spanning a maximum of 1.32°C for probe pair BP1_3. GC content of all probes in the final IGF set was also calculated. Mean inter-probe GC% was determined to be 48.96%. This spanned a 48.87% range from 29.70–78.57% (BP3_14C and BP3_6T respectively), while mean intra-probe GC% is 5.25%, spanning a maximum of 17.77% for probe pair BP3_3.

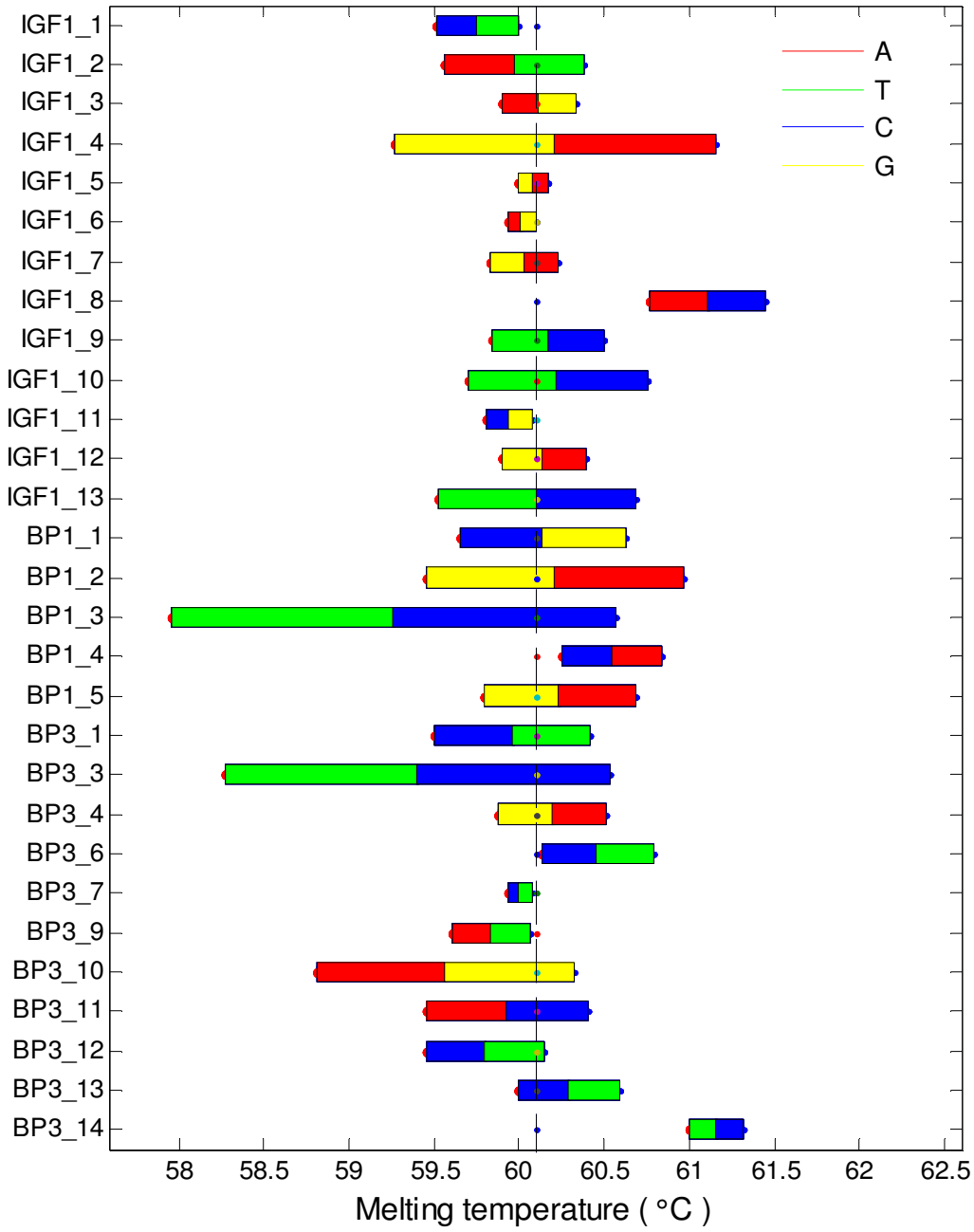


Figure 19: Primer3 (Panjokovich consensus) probe predicted T_M. The predicted T_M for each probe pair is displayed on a single horizontal axis with box colour representative of the tagSNP allele. The vertical terminal box positions denote the predicted optimal T_M for each probe. The intra-probe vertical line separating probe pairs denotes the predicted optimal T_M for that pair. The central dotted vertical line denotes that optimal predicted T_M for all probes in the full IGF set. Mean inter-primer T_M is 60.09°C, spanning 3.49°C from 57.96 to 61.45°C (BP1_3C and IGF1_8A respectively), while mean intra-primer T_M is 0.548 °C, spanning a maximum of 1.32°C (for probe pair BP1_3).

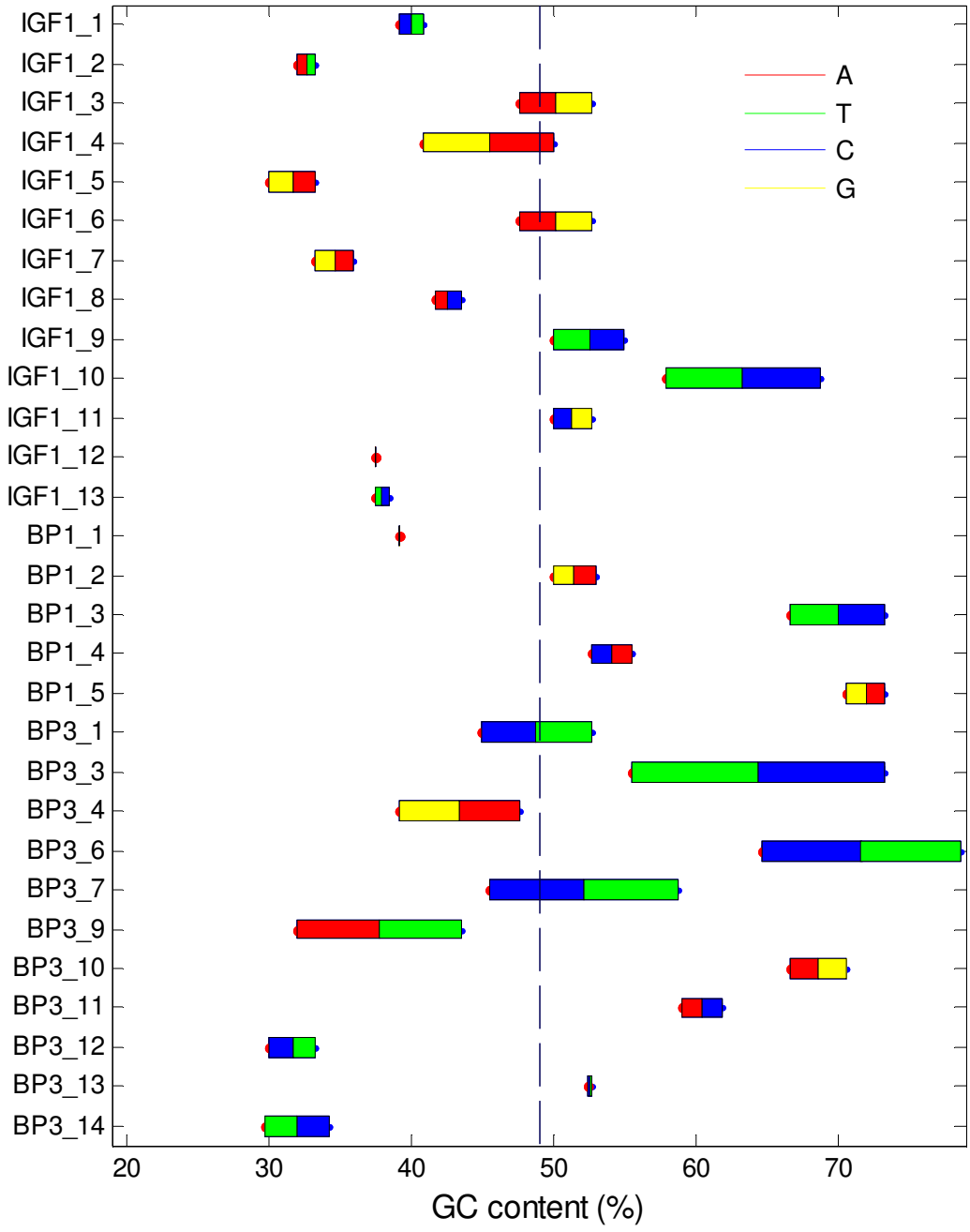


Figure 20: Probe GC content. GC content of all probes in the final IGF set are shown. GC content for each probe is displayed with box colour representative of the tagSNP allele. The vertical terminal box position denotes the GC % for each probe while the intra-primer vertical line separating probe pairs denotes the mean GC% for that pair. The central dotted vertical line denotes the mean GC% for all probes in the full IGF set. Mean inter-probe GC% is 48.96%, spanning 48.87% from 29.70–78.57% (BP3_14C and BP3_6T respectively), while mean intra-probe GC% is 5.25%, spanning a maximum of 17.77% for probe pair BP3_3.

3.3.5 UCSC PCR

UCSC PCR was used to investigate primer specificity; like blast searching this program compares primer pairs against a specified genomic repository (in this case the Human Genome March 2006 assembly). Sequences are mapped to their respective genomic loci and intervening sequence regions returned to allow primer pair specificity and amplicon size predictions to be made. Figure 21 and Figure 22 display the output received for primer pairs IGF1_5 and IGF1_9 respectively. IGF1_9 primer pairs failed to produce an amplicon while primer pair IGF1_5 (like all other IGF multiplex set primer pairs) produced a single target sequence of the anticipated size. All predicted amplicon sequences are displayed in electronic appendix Table 6.

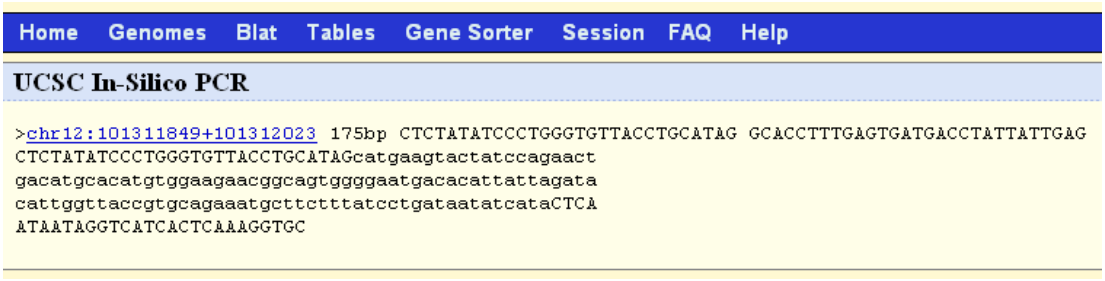


Figure 21; UCSC *in silico* PCR amplification; successful PCR amplification was predicted for 26 of the 27 primer pairs tested using this approach. The example displayed shows IGF1_5 primers and single putative 175bp amplicon spanning IGF1 chromosome 12:101311849+101312023.

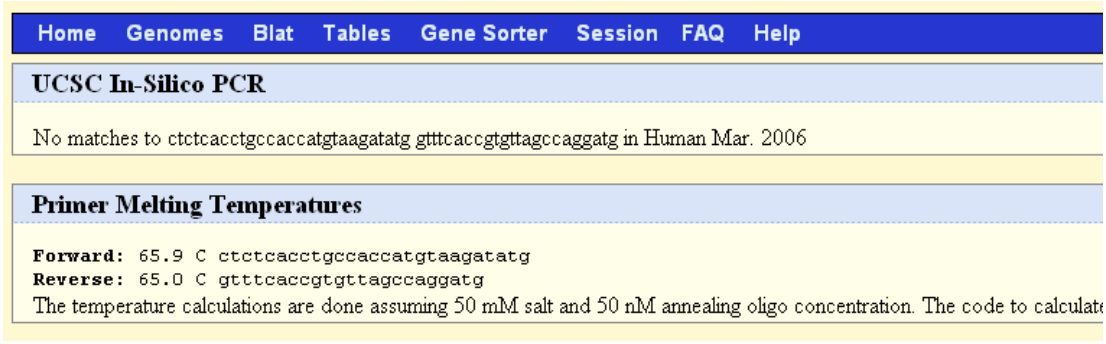


Figure 22: UCSC *in silico* PCR amplification failed to produce an amplicon for IGF1_9 primer pair within either default or relaxed minimum perfect match settings of 15 or 1 respectively.

3.3.6 UCSC Amplicon Analysis

UCSC Genome View was used to access information regarding constituent polymorphism indels and low complexity repeat regions which may affect amplicon size, the number of amplicons produced and efficiency / specificity of PCR amplification or probe hybridisation. UCSC repeat region determination was made for IGF1_9, IGF1_10 and BP3_14 targets (as displayed in Figure 23, Figure 24 and Figure 25 respectively). IGF1_9 contains one LTR (long terminal repeat) and one SINE (short internuclear dispersed) region and no indels. IGF_10 contains one SINE and one partially overlapping simple repeat region as well as two indel polymorphisms (rs57468885 and rs58322331) seven and one nucleotides in length. BP3_14 was also found to contain two LTRs and one simple sequence/tandem repeat region. This target also highly polymorphic containing five indel polymorphisms (rs34122177, rs55702604, rs35919935, rs58209457 and rs34087654) of four, three, two, five and five nucleotides respectively.

The polymorphic nature of BP1_3/4 and BP3_10/11 targets are also displayed (Figure 26 and Figure 27). BP1_3/4 primer pair target region. UCSC Genome View was used to access information regarding amplicon-constituent polymorphisms. BP1_3/4 amplifies two target tagSNPs rs3828998 and rs9658194 (IGFBP1 bins 3 and 4 respectively) and one intervening SNP rs9658195. BP3_10/11 primer pair amplifies two target tagSNPs rs35751739 and rs35496550 (IGFBP3 bins 10 and 11 respectively) and one intervening SNP rs34735423.

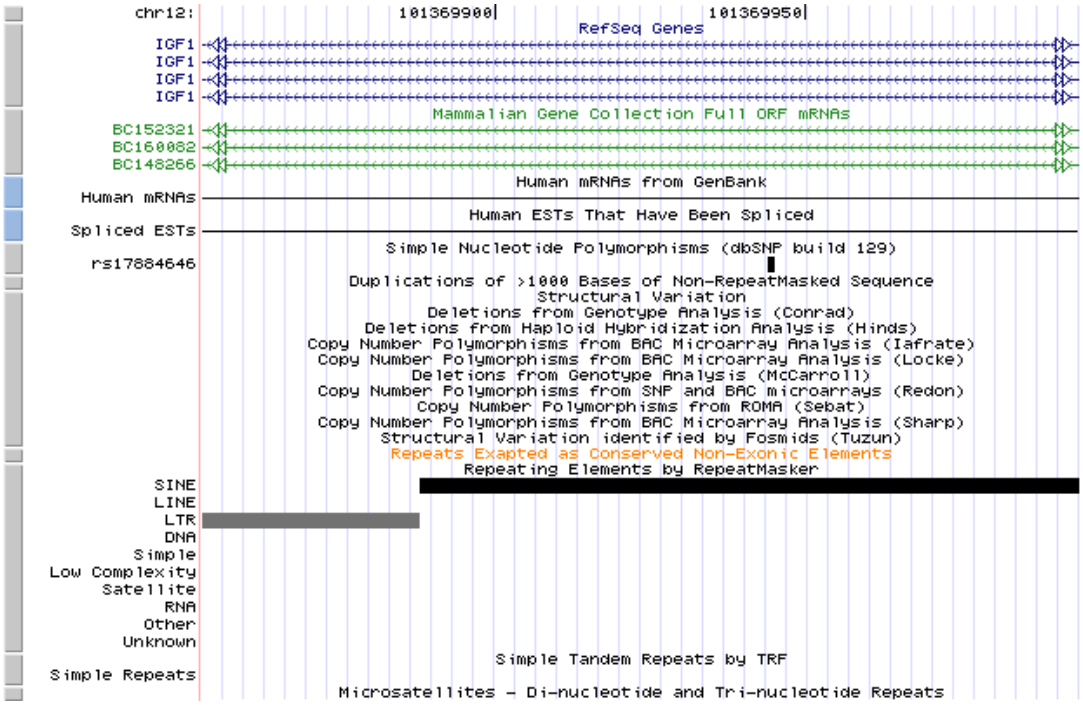


Figure 23: IGF1_9 primer pair target region. UCSC Genome View was used to access information regarding constituent polymorphisms and repeat regions which may affect amplicon size, the number of amplicons produced or the efficiency / specificity of PCR amplification. No polymorphism indels or repeat regions were identified however LTR (long terminal repeat) and SINE (short internuclear dispersed) regions were identified.

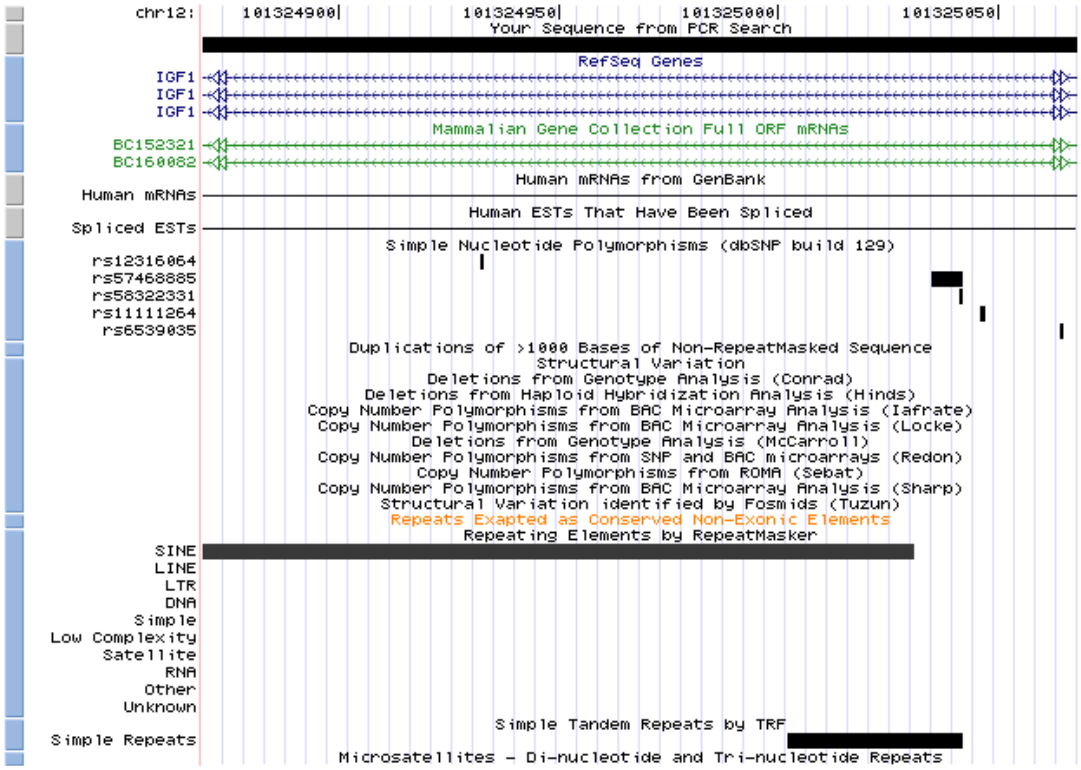


Figure 24: IGF1_10 primer pair target region. UCSC Genome View was used to access information regarding constituent polymorphisms and repeat regions which may affect amplicon size, the number of amplicons produced or the efficiency/specificity of PCR amplification. Two indel polymorphisms (rs57468885 and rs58322331), one SINE (short internuclear dispersed region) and one partially overlapping simple repeat region are evident.

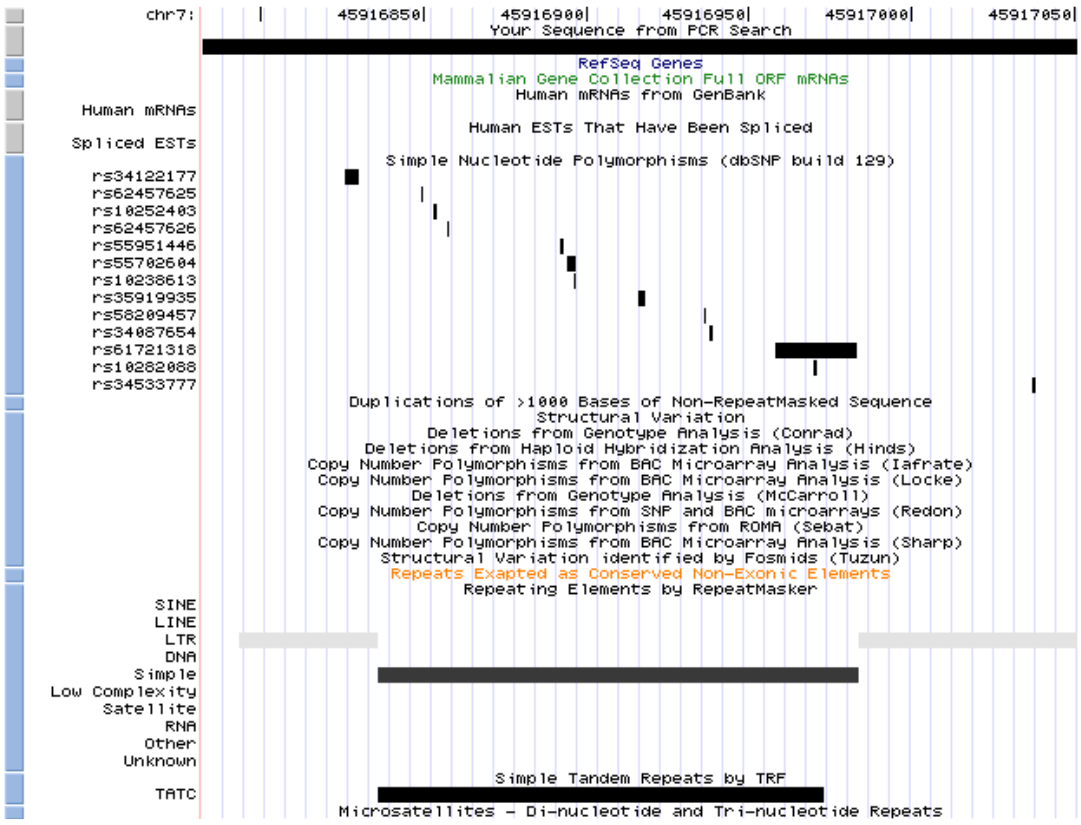


Figure 25: BP3_14 primer pair target region. UCSC Genome View was used to access information regarding constituent polymorphisms and repeat regions which may affect amplicon size, the number of amplicons produced or the efficiency/specificity of PCR amplification. Two LTR (long terminal repeat), one simple sequence/tandem repeat region were identified. Five indel polymorphisms (rs34122177, rs55702604, rs35919935, rs58209457 and rs34087654) are also evident.

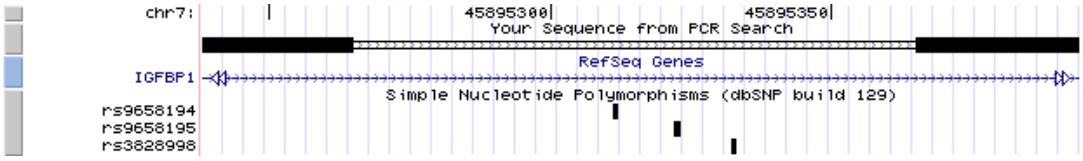


Figure 26: BP1_3/4 primer pair target region. UCSC Genome View was used to access information regarding amplicon-constituent polymorphisms. Two target tagSNPs rs3828998 and rs9658194, IGFBP1 bins 3 and 4 respectively are amplified with intervening SNP rs9658195.

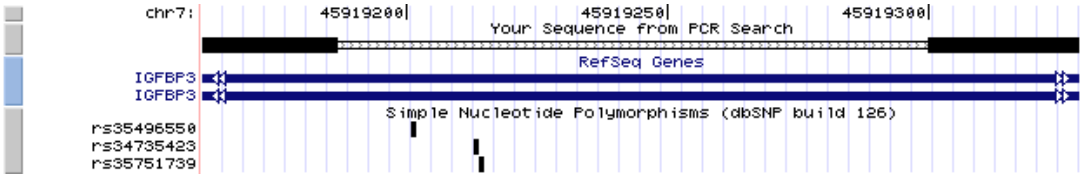


Figure 27: BP3_10/11 primer pair target region. UCSC Genome View was used to access information regarding amplicon-constituent polymorphisms. Two target tagSNPs rs35751739 and rs35496550, IGFBP3 bins 10 and 11 respectively are amplified with intervening SNP rs34735423.

Allele frequency data for both intervening SNPs rs9658195 and rs34735423 were determined using dbSNP. SNP rs9658195 displays a minor allele frequency of 0.011 in PDR90 mixed population while SNP rs34735423 does not exhibit allele G in European population EGP_CEPH-PANEL. The position and allele frequency of these intervening SNPs was considered to facilitate robust probe design.

Table 15: Allele Frequency data for intervening SNPs rs9658195 and rs34735423. Reference sequence identifiers, population and allele frequencies are displayed. SNP rs9658195 displays a minor allele frequency of 0.011 in PDR90 mixed population while SNP rs34735423 does not exhibit allele G in European population EGP_CEPH-PANEL.

refSNP ID	Population	Allele Frequency	
		C	T
rs9658195	PDR90	0.989	0.011
		A	**G**
rs34735423	EGP_CEPH-PANEL	**0.00**	1.000

3.3.7 Primer Pair Proximity analysis

Primer Map was used to position all primer pair sequences across their respective genes. The visual output allows identification of overlapping primer pairs likely to induce amplification failure or those in close proximity, which may facilitate nonspecific amplification between adjacent non-pair forward and reverse primers. Overlapping pairs were not admissible and were redesigned or removed. Primer pairs producing nonspecific products of <550bp also received this treatment while those in the 550-800bp range were noted to facilitate thermal cycle adjustment / optimisation in subsequent reactions. Three nonspecific products of 583, 589 and 642 within this 550-800bp range were predicted for adjacent non-pair primers IGF1_2F/IGF1_8R, BP3_12F/BP3_10/11R and BP3_7F/BP3_1R as displayed in Figure 28, Figure 29 and Figure 30 respectively.

Figure 28: Gene Infinity Primer Map Output. Both primer pairs IGF1_2 and IGF1_8 are in close genomic proximity; as such a nonspecific product of 583bp may be produced between IGF1_2F and IGF1_8R (forward and reverse primers of alternate pairs). Forward and reverse primer loci are highlighted in pink and orange respectively, while red arrows pinpoint primer pair target sequences.

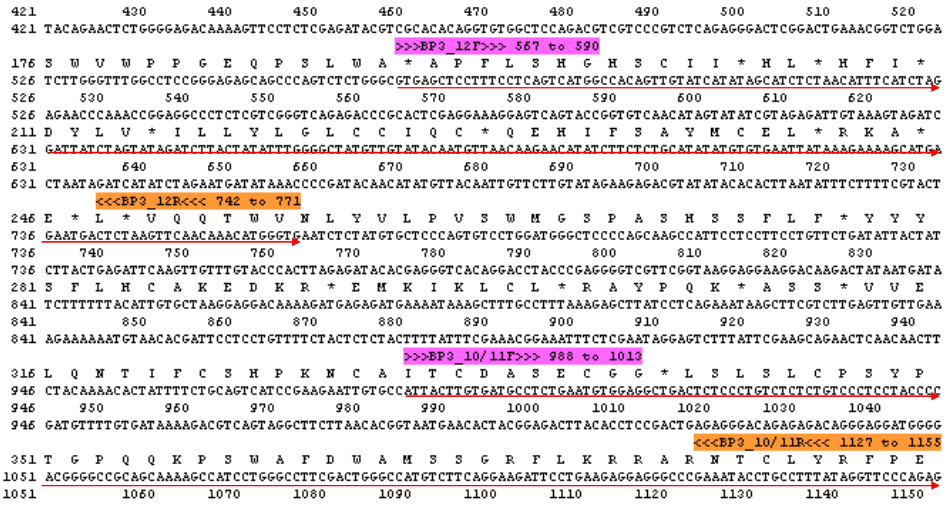


Figure 29: Gene Infinity Primer Map Output. Both primer pairs BP3_12 and BP3_10/11 are in close genomic proximity; as such a nonspecific product of 589bp may be produced between BP3_12F and BP3_10/11R (forward and reverse primers of alternate pairs). Forward and reverse primer loci are highlighted in pink and orange respectively, while red arrows pinpoint primer pair target sequences.

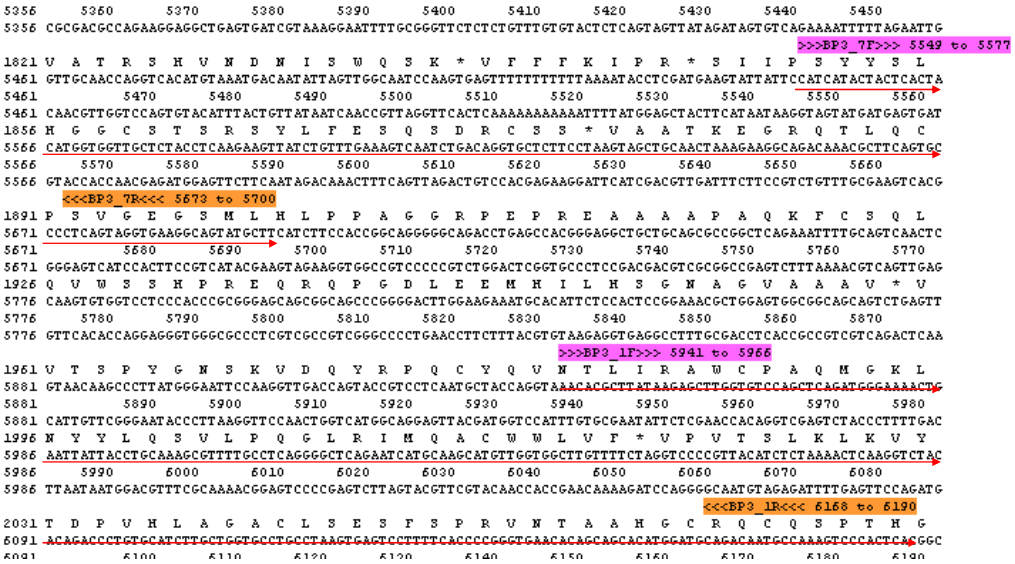


Figure 30: Gene Infinity Primer Map Output. Both primer pairs BP3_7 and BP3_1 are in close genomic proximity; as such a nonspecific product of 642bp may be produced between BP3_7F and BP3_1R (forward and reverse primers of alternate pairs). Forward and reverse primer loci are highlighted in pink and orange respectively, while red arrows pinpoint primer pair target sequences.

3.3.8 AutoDimer Cross-homology

AutoDimer was used to analyse all primers with regard to their propensity towards homodimer, heterodimer and hairpin formation. Primer dimers displaying potentially problematic score thresholds of >7 were redesigned or excluded from analysis if no suitable alternative could be identified. Dimers receiving threshold scores of 7 were further analysed in terms of their putative stability; duplexes receiving predicted T_Ms of <20°C and Gibbs folding free energy (delta G) >0 kcal/mol (at 64°C) were deemed permissible as these are unlikely to hold their structure given the lowest PCR reaction temperature (-1°C). A single hetero-dimer receiving a score of 7 and delta G >0 kcal/mol (at 64°C) was included in the final IGF multiplex set (see Figure 31).

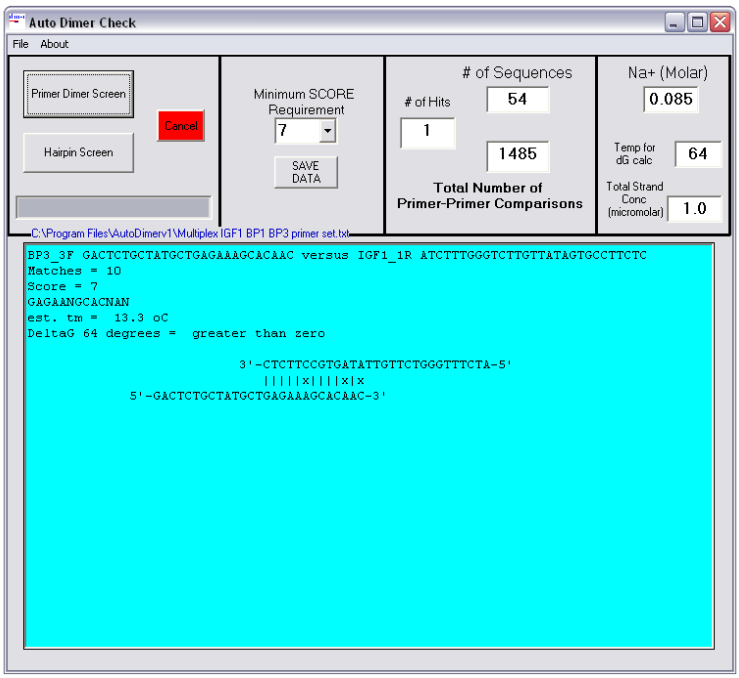


Figure 31: AutoDimer primer dimer results for final multiplex set. A single potentially problematic heterodimer between IGF1_1R and BP3_3F was predicted. Both primers display complementarity at 3' ends; a feature which tends to promote primer dimer formation. A borderline score of 7 was also calculated however its thermodynamic profile indicates that primer dimer formation is unlikely to be problematic given the PCR conditions under which it is to be amplified (delta G >0 kcal/mol). No potentially problematic homodimer or hairpin structures were identified.

3.4 Discussion

Extensive evaluation of primer pairs and probes was carried out to maximise the probability of successful high-dimensionality multiplex target amplification and downstream SAT genotyping. The following section describes the results of PCR primer and probe design, *in silico* analysis and theoretical bin coverage achieved using the approaches described.

Using Primer3, primer pairs for all tagSNPs generated by Tagger were designed within strict parameter thresholds in terms of inter- and intra-Tm, GC%, amplicon length etc. Pairs were analysed for specificity using UCSC PCR and evaluated in terms of their constituent sequence type and insertion / deletion polymorphisms, which may affect amplification profiles or probe hybridisation. Proximity analysis was performed using Primer Map and multiplex compatibility of primer pairs and probes was assessed using AutoDimer until a primer pair / probe set representing the largest number of bins across candidate genes IGF1, IGFBP1 and IGFBP3 was identified. The results of primer / probe design and *in silico* evaluation techniques employed during this important design phase are described.

3.4.1 PCR

3.4.1.1 Primer Profiles (Annealing Temperature, GC Content and Distribution)

Following nearest neighbour-based primer T_M approximation it is recommended that primers are tested empirically and amended to suit the true primer pair requirements. This approach is unsuitable for multiplex PCR however where one highly specific T_M is required to meet the needs of a full primer set. In such cases theoretical anomalies must be minimized before ordering primer sets [141]. Work has been carried out by a number of groups in an attempt to develop a method which accurately and precisely quantifies base stacking forces, however calculation methods for alternate algorithms result in small but significant differences in predicted T_M [156,157]. Panjkovich *et al*., have devised a consensus T_M calculation method which employs multiple nearest neighbour inclusive formulas to derive mean T_M scores with minimal error probabilities [141]. This method has been implemented in Primer3 (primer design program) for T_A prediction and has been used to successfully design a number of large multiplexes including a 10-plex, 25-plex and 52-plex for forensic applications [133,158,158,159].

Using Primer3 primer design software, 27 primer pairs spanning 29 target bins were designed. These displayed an average primer length of 27 bases ranging from 22-30 nucleotides, corresponding to a mean Panjkovich derived T_A of 65.25 ^oC. An inter-primer T_A of just 1.68 ^oC (64.31-65.99 ^oC) was predicted across the full primer set, with a maximum intra-primer pair T_A of 1.32^oC predicted for BP3_10/11 primer pair. A number of other successful multiplex studies exceeding the T_A ranges predicted for this IGF multiplex set have been performed successfully with Sanchez *et al*., successfully using

intra /inter primer T_A ranges of 4.0 and 7.0°C respectively for multiplex amplification of 52 targets [134,141,160,161]. As such T_A divergence is unlikely to effect multiplex amplification in this instance. Primer lengths and associated T_A predicted for this primer set exceed those usually seen for singleplex PCR, however a number of groups including Henegariu *et al*., and Dieffenbach *et al*., have recently used primers in this range for high-specificity multiplex amplification with excellent results [132,133,162].

Alternate bond strengths contributed by guanine-cytosine (GC) triple and adenine-thymine (AT) double bonds can impact hybridization characteristics of PCR primers, with higher AT sequences being less stable than their higher GC counterparts. Nowhere is this more critical than at 3' terminal ends. Polymerase effects elongation by addition of dNTPs to 3' sticky ends. Use of GC clamps at these terminal positions have been found to increase duplex stability, specificity and efficiency of amplification and as such a GC clamping provision was included in the design specification [135-138]. All primers in the final set adhered to this specification bar IGF1_10 which contains a weak "A" and "TTT" terminal nucleotide sequence for forward and reverse primer pairs respectively. Loss of this GC clamping feature may affect functionality of this primer pair.

It is also recommended that GC content for all primers be maintained between 40-60%. All primers in the final set adhered to this ideal specification with mean inter-primer GC of 46.31%, spanning 15% from 40.0–55.0% for primers IGF1_6R/BP3_12R and BP3_6R respectively. A mean intra-primer GC% of 4.33% was also calculated spanning a maximum of 14% for BP3_12. This is still reasonably narrow and should not affect amplification efficiency to a significant extent.

3.4.1.2 Primer Specificity

UCSC PCR was used to investigate primer specificity; this program compares inputted nucleotide primer sequences against the specified genomic repository (in this case the Human Genome March 2006 assembly). This acts to map primer sequences to their respective genomic loci and returns both primers and intervening sequence regions allowing primer pair specificity and amplicon size to be determined. *In silico* PCR analysis predicted target–specific amplification with amplicons in the anticipated location and of the anticipated size range for all IGF multiplex set primer pairs bar IGF1_9 which failed to produce a product. The reason for this failure is unknown however IGF1_9 is located in a repeat region where sequence determination is often less concrete, as such differences are more likely to be observed in this type of genomic environment. Using the Ensembl-derived rs17884646 flanking sequence; IGF1_9 primers were predicted to produce a 141 base amplicon and this primer pair was included in the final IGF multiplex set.

3.4.1.3 PCR Amplicon length

Another key consideration for multiplex PCR analysis is amplicon length; singleplex PCR amplification has been successfully performed on targets ranging in size from 44bp to 27 kb [163,164]. However optimal reaction components, concentrations and thermocycle segment times and efficiencies differ significantly between alternately sized amplicons. Extension time requirements are particularly increased with longer amplicon lengths (i.e. using *Taq* polymerase with processivity of 60nt/sec, theoretical extension

times of 5, 16 and 166 seconds are required for amplification of 300bp, 1kb and 10kb amplicons respectively). Application of single extension times to multiplex amplification can result in skewed relative yields for alternately sized amplicons, it was therefore important to ensure the amplicon size range was kept relatively narrow.

The issue of amplicon length with regard to its potential influence on downstream SAT hybridization efficiency and specificity was also considered. A recent study by Liu *et al*., used a range of targets of alternate lengths, ranging in size from 1490-93bp, to assess the effect of length on hybridisation and efficiency. Hybridisation efficiency was found to increase with decreasing target length (by up to a factor of 8.8 for smaller 93-145bp amplicons) and was also found to reduce the incidence of false negatives. Shorter fragments were found to marginally increase false positive error rates however, thus a compromise between hybridisation efficiency and specificity is required [165-168]. A number of groups including Dunbar and Armstrong *et al*., have used a range of amplicon target sizes from 100-300bp for SAT analysis with good effect [10,150,153,169]. An optimal range of 100 – 300bp was therefore set for this IGF multiplex set. A total of 25 out of 27 amplicons fell within this range, although tagSNP rs3730204 (IGF1_12) flanking sequence contains a disproportionately high distribution of AT; as such a larger allowable threshold was set and primer pairs spanning 500bp produced. Rs1019731 (IGF1_8) also contained a disproportionate GC/AT distribution and a smaller allowable amplicon size was therefore accepted (72bp). A mean multiplex amplicon length of 209bp with standard deviation, variance and range of 78, 6097 and 428bp respectively was determined for all putative amplicons in the final IGF Multiplex set.

3.4.1.4 Primer Targets

Target sequence structures can impact specificity and efficiency of amplification, as such target sequences were assessed using UCSC Genome Browser detailed view as described. Of particular interest was identification of repeat region sequences. Successful amplification of repeat region targets is notoriously difficult; multiple repeats often lie in close proximity within PCR amplification range, resulting in amplification of longer products. Additionally the repetitive nature of this sequence type also lends itself to secondary structure (loop) formation which can facilitate deletion mutagenesis and production of shorter PCR products [170]. Despite this, repeat regions have been successfully amplified in the past and therefore, where representation of these regions could not be achieved by use of tagSNPs spanning more PCR-facilitations environments, repeat region primer pairs were designed and included in the IGF multiplex set. Primer pairs IGF1_9, IGF1_10 and BP3_14 fall in this category, spanning Alu-SINE/LTR, Alu-SINE/ Simple repeat and LTR/Simple tandem repeat regions respectively. Alu-SINE targets were of particular interest as these types of retrotransposon have been found by Jiang *et al*., to induce hypermethylation of the Gck promoter, thereby reducing hepatic expression and elevating diabetogenic potential in ageing rats [171-173].

Inclusion of repeat region targets can also add an extra dimension of complexity due to their highly polymorphic nature which may include indels. Due to the homologous nature of chromosomes, sequence length differences caused by insertion/deletion polymorphisms in hereditary alleles may result in production of two amplicons of alternate size from high specificity amplification. IGF1_9 amplicon does not contain any known indels, however IGF1_10 contains two (rs57468885, rs58322331) and BP3_14

contains five indels (rs34122177, rs55702604, rs35919935, rs58209457 and rs34087654) of 7, 25 and 4, 3, 2, 5, 5 nucleotides respectively. Thus amplicons may range from 192-224bp and 274- 293bp for IGF1_10 and BP3_14 primer pairs respectively.

3.4.1.5 Primer Pair Proximity Analysis

Gene Infinity's Primer Map program was used to position all primer pair sequences across their respective genes. Primers amplifying overlapping targets may result in amplification of short products, poor target amplification, or amplification failure while non-overlapping amplicons located in close genomic proximity may allow amplification of longer unspecific between adjacent non-pair forward and reverse primers. Longer PCR products have different thermal cycle and reaction requirements however and as such long non-specific products which significantly exceed the maximal target size are unlikely to be amplified within optimised reactions [174]. Smaller nonspecific amplicons may cause problems however and as such those within 300bp of the largest target-specific amplicon were identified. Three potential nonspecific products of 583, 589 and 642bp within this 550-800bp range were predicted for adjacent non-pair primers IGF1_2F/IGF1_8R, BP3_12F/BP3_10/11R and BP3_7F/BP3_1R respectively. Thermal cycle and protocol adjustments may be required to remove these nonspecific products in final multiplex amplifications [162].

3.4.1.6 Primer Cross-Reactivity

Primer3 has limitations in terms of its ability to predict multiplex compatibility. The propensity of primers toward homodimer, heterodimer and hairpin formation is particularly significant for multiplex amplification where complexity is increased. A number of programs including FastPCR and Beacon Designer facilitate analysis of this type during the primer design phase; however FastPCR uses the Allawi and SantaLucia nearest-neighbour algorithm for T_M calculation, and Beacon Designer requires licence payment and limits simultaneous evaluation to five primer pairs [175]. As such, an independent post-primer design analysis program "AutoDimer" was used for evaluation of Primer3-designed primers.

A score threshold of seven, as recommend by AutoDimer, was used to identify potentially problematic primers. Primers displaying a high propensity towards homodimer and heterodimer formation were redesigned where possible and replaced by lower scoring alternatives. IGFBP3 bins two, five and eight (SNP rs33979592 and self tagging SNPs rs2854744 and rs2453840) were excluded from multiplex inclusion due to high primer dimer scores (>13) received for all possible primers in these regions. None of these SNPs were predicted to confer a deleterious impact during PupaSuite functional impact prediction (chapter 2). The AutoDimer sliding algorithm was used to perform 1,485 primer-primer comparisons[8] on the final 54-primer set. Just two of these included in the final IGF multiplex set; BP3_3F and IGF1_1R, displayed a potentially problematic

[8] The number of duplex comparisons made may be calculated using the following formula $2n^2 + n$, where n = the number of primer pairs

hetero-dimer score of seven. Ten complementary matches spanning thirteen nucleotides near both 3' ends of this duplex were predicted; this is potentially significant as polymerase effects elongation by addition of dNTPs to 3'sticky ends. IGF1_1R also contains a terminal guanine clamp, which may further act to stabilise the BP3_3F/IGF1_1R duplex [136,176]. Gibbs folding free energy can be used to determine the thermal profile of a molecule within a closed system (i.e. PCR reaction), as such the minimum putative temperature used during PCR amplification was applied prior to analysis. A delta G calculation threshold temperature of 64 rather than the predicted optimal 65°C was used, this was to facilitate thermal cycle block and / or optimal T_M prediction error. A non-significant delta G of >0 kcal/mole was predicted, indicating that the duplex is likely to be unstable under given PCR conditions. As such, both primers were included in the final multiplex set [176,177]. A hairpin screen, using an AutoDimer threshold score of seven was also performed, however problematic primers of this nature were not found to be present [176]. Primer3 implements an intra-primer and hairpin score threshold of eight and therefore self and intra-primer homology was protected to a large extent during primer design.

3.4.2 SAT

3.4.2.1 Probe Profile (Annealing Temperature, GC Content and SNP position)

In order to achieve maximum hybridisation signal strength and discrimination it is essential for all probe sequences to have a narrow T_M range. Again Primer3's nearest-neighbour calculation was implemented to design probes and predict associated T_Ms. A mean T_M of 60.09°C for all IGF set probes was predicted, ranging 3.49°C from 57.96 to

61.45°C for BP1_3C and IGF1_8A respectively. Maximum intra-probe pair T_M was predicted to be 2.61°C for BP1_3. This mirrors a narrow T_M range used by Xu *et al*., who successfully used maximum intra-probe T_M range of 4.8°C from 62.1-66.9°C to successfully perform SAT SNP genotyping [10].

Probe lengths are of less concern for this type of analysis relative to PCR-based primer design, due to the limited capacity for cross-homology incurred by use of PCR targets. All our probe sequences fell within 22 bases, ranging from 14 bases for BP3_6T to 37 bases for BP3_14C. A range of 15-25 bases is commonplace for SAT-based genotyping assays; however, sequences of 100 bases have been used for microarray analysis and therefore it is anticipated that our largest sequence should function suitably [10,152,153].

Letowski *et al*., determined that centrally positioned mismatches contribute to greater duplex instability than those located at non-central positions [152]. As such, an attempt was made to position SNPs centrally within probe sequences. This was not possible for all targets however, due to sequence secondary structure and adjacent SNP constraints and as such, a number of alleles were positioned non-centrally. A somewhat arbitrary designation of centrality was applied to classify probes with regard to this feature, with SNPs positioned <25% from probe terminus designated as skewed. Four probe pairs including IGF1_2, IGF1_8, IGF1_9 and IGF1_11 fell within this category and may display reduced discrimination capacity.

Broad GC ranges of ~25-70% are permissible for hybridisation probes, with lower GC probes displaying reduced hybridisation signals, and those exceeding 80% displaying a stronger propensity for cross-homology [10,153,154,178]. The mean GC content for IGF multiplex set probes was determined to be 48.96% ranging from 29.70-78.57% for BP3_14C and BP3_6T respectively. Inter-probe pair GC deviance within the allowable range should be of little consequence due to the comparative nature of intra-probe pair allele designation, although it is hypothesised that intra-probe pair GC divergence may contribute some effect. A number of probe pairs including IGF1_4, IGF1_10, BP3_9, BP3_7, BP3_6 and BP3_3T in our final IGF multiplex set exceed the ~7°C intra-probe pair GC divergence seen in previous SAT genotyping assays. The consequence of this is unknown. T_M similarity within probe pairs may be sufficient to facilitate accurate allele calling, alternatively however, T_M restrictions may need to be relaxed to facilitate higher GC similarity within the pair if experimental evaluation determines this feature to be problematic [153].

3.4.2.2 Probe Proximity Analysis

The proximity of adjacent tagSNPs was determined using the UCSC Genome Browser as described. TagSNPs <30 bases apart that could not be substituted for alternate tags were amplified within one sequence using single primer pairs. As such these pairs will require double amplification (in separate reactions) to allow downstream SAT genotyping. Two such pairs; BP1_3/4 and BP3_10/11 were included in the final IGF multiplex set. Primer pair BP1_3/4 contains two tagSNPs rs3828998 and rs9658194 (IGFBP1 bins 3 and 4 respectively) just 21 bases apart, while primer pair BP3_10/11 contains tagSNPs

rs35751739 and rs35496550 (IGFBP3 bins 10 and 11 respectively) just 13 bases apart. To complicate matters further, both contained a third variation positioned between target tagSNPs. In both cases allele frequency data and nature of the intervening third variation was assessed. SNP rs34735423 (BP3_10/11 intervening SNP) is thought to be absent in European Caucasian derived populations (EGP CEPH-panel) and is therefore also very unlikely to cause complication in terms of downstream probe hybridization. SNP rs9658195 (BP1_3/4 intervening SNP), has a relatively small minor allele frequency of 0.011 and should also cause minimal disruption. PDR90 global mixed population genotype data only was available for this SNP however; therefore European Caucasian population specific genotyping of this locus may act to further clarify this matter.

3.4.2.3 Probe Cross-Reactivity

AutoDimer was again used for sequence cross-reactivity analysis. No problematic duplexes were predicted for probes in the final multiplex set, cross-reactivity should therefore not be an issue using the probe set described [142].

The work in this chapter aimed to use a range of *in silico* design and analysis programs to increase the probability of high experimental multiplex PCR and SAT functionality for maximum coverage of tagSNP target bins (as describe chapter 2). As described in our project objectives we aimed to *" perform multiplex primer and probe design, utilising in silico and manual analysis for evaluation and selection of a high specificity primer / probe sets in a manner compatible with downstream multiplex PCR and SAT analysis"*. This was achieved by testing the applicability of all tagSNPs using a range of *in silico*

techniques. Primers and probes were designed within strict specifications which should facilitate high-functionality. A number of concessions regarding primer design were made to allow inclusion of primers which displayed sub-optimal profiles; several primers which

displayed very poor profiles were excluded from further analysis (IGFBP3 bins two, five and eight). This allowed design of a 27 primer- / 29 probe- pair set which theoretically allows representation of 29 bins from the original 32 bins identified by Haploview. Experimental analysis of these sequences is required however to determine their true functionality.

The following chapter aims to analyse all primer sequences experimentally; testing specificity and multiplex compatibility for construction of a number of well-defined multiplex sets which facilitate robust amplification of IGF target sequences.

Chapter 4: Multiplex polymerase chain reaction

4.1 Overview

Chapter 3 *primer / probe design and in silico evaluation* aimed to design a set of multiplex-compatible PCR primer and SAT probe sequences spanning the maximum number of tagSNP bins in the candidate region. Careful primer design and *in silico* evaluation was used to design these sequences for target-specific, high-yield amplification. However while *in silico* evaluation of this nature can act to improve the probability of high-level primer performance, it is no guarantee of experimental success. As such, chapter 4 aims to evaluate primer pair in terms of their true experimental functionality with regards to both singleplex and multiplex amplification. Multiplex critical primer pair parameters are investigated extensively and results discussed with respect to theoretically derived profiles. Singleplex profiles are also compared to those derived during multiplex amplification. This chapter also provides an overview of reaction components which may be used to ameliorate PCR amplification and a number of these techniques are implemented to attempt to improve amplification of repeat region targets.

4.1.1 Experimental PCR Considerations

An understanding of PCR constituents; their function, scope and inter-relationships within a given reaction, can act to increase specificity, efficiency of amplification and reduce PCR optimisation times. The influence of magnesium chloride ($MgCl_2$), dNTP, primer and template concentrations as well as adjuvant addition and polymerases selection are considered.

It is recommended that primer concentration be set with a molar excess of ~10^7 with respect to template concentration if non-specific artefact formation and preferential amplification of GC rich targets is to be avoided [179]. Preferential amplification of targets is thought to occur by two basic mechanisms; PCR drift or inhibition. PCR drift occurs as a direct result of inadequate template concentrations that cause fluctuations in reagent interaction and preferential and non-specific amplification, while selection is induced by intrinsic template properties, namely divergent GC contents or structural anomalies which affect amplification efficiency [180].

Increased total template concentrations may be used to diminish the effects of drift, while increased relative primer concentrations can act to offset less efficient amplification of poorly amplified targets in multiplex sets. This strategy does not always work for repeat region targets prone to secondary structure formation however; in such instances adjuvant addition may be required [181]. A wide range of adjuvants including single-stranded DNA binding proteins (gp32 and EcoSSB), non-ionic detergents (Tween-20, Nonidet P-40 and Triton X-100) and organic solvents (formamide and Dimethyl sulfoxide) have been developed. These employ an array of mechanisms to relax hairpin structures while allowing conditions conducive to hybridisation to be retained [182].

Taq DNA polymerase is a magnesium-dependant enzyme and as such, sufficient magnesium chloride addition is required if high-yields are to be produced. Divalent cations act to stabilise hydrogen bond formation between single-stranded nucleotides and nucleic acid sequences. As a result dNTP, template and primer concentrations all influence Mg^{2+} availability. Mg^{2+} increases duplex stability however if used in excess it

can result in poor dsDNA melting during denaturation. Excessive Mg^{2+} concentrations have been shown to cause significant problems for amplification of repeat-region targets; where magnesium acts to stabilise hairpins facilitating non-specific amplification of aberrant targets and poor target yields [162,180,183].

The type of polymerase selected is also an important consideration. An array of modified polymerases have been developed to improve reaction speeds, fidelity and target-specific amplification. Hot-start, 3'-5' exonuclease-enabled, high-processivity, high-fidelity and inhibition-resistant polymerases are outlined in the following section.

Low reaction temperatures incurred during PCR preparation or thermocycle ramping can allow non-specific inter- and intra-primer binding even for those with non-significant designations[9]. By implementing hot-start procedures with limited polymerase activity prior to PCR cycling, non-specific amplification can be reduced and target-specific yields increased. These procedures were originally performed by physical removal or barrier-facilitated sequestration of reaction essential components (i.e. polymerase), however more recent innovations employ modified polymerases with chemical or ligand-mediated (oligonucleotide or antibody) active-site inactivation with good effect [184-186]. The choice between ligand- and chemically-inactivated polymerase depends on the application; ligand-mediated moieties require very short temperature activation and as such, retain higher polymerase activity throughout the reaction. Chemically modified polymerases however, can be used to reduce misamplification of problematic targets by time-release activation. This technique uses an incomplete preliminary activation step

[9] As determined by cross-reactivity prediction software (i.e. AutoDimer)

allowing polymerase activation over a number of cycles. Consequently, active polymerase concentrations mirror target availability and excess polymerase induced non-specific amplification may be reduced [186,187].

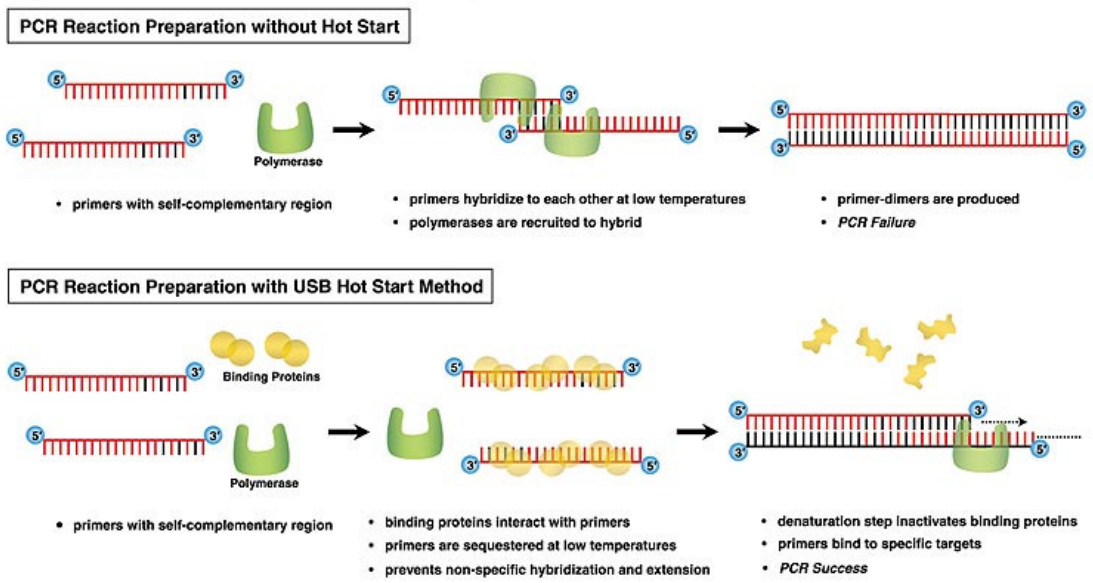


Figure 32: Hot start PCR. The mechanism by which primer dimer formation occurs during non-hot start PCR amplification is displayed as is binding protein mediated hot start amplification. Using the hot start method described; binding proteins bind with single stranded primers at low temperatures thereby preventing non-specific primer-primer binding and subsequent primer dimer formation [188].

Polymerase selection is also known to affect the fidelity of PCR amplification. In fact it is thought that polymerase may be responsible for the majority of errors that arise in the form of misincorporation substitutions [170]. A number of aspects impact on the fidelatous capacity of the enzyme including; the specificity of dNTP binding, rate of phosphodiester bond formation, pyrophosphate release, extension following misincorporation and the ability of the enzyme to carryout proofreading or 3'-5' exonuclease excision of erroneously incorporated nucleotides following aberrant inclusion. This proofreading feature is illustrated in figure Figure 33 [170,189-192].

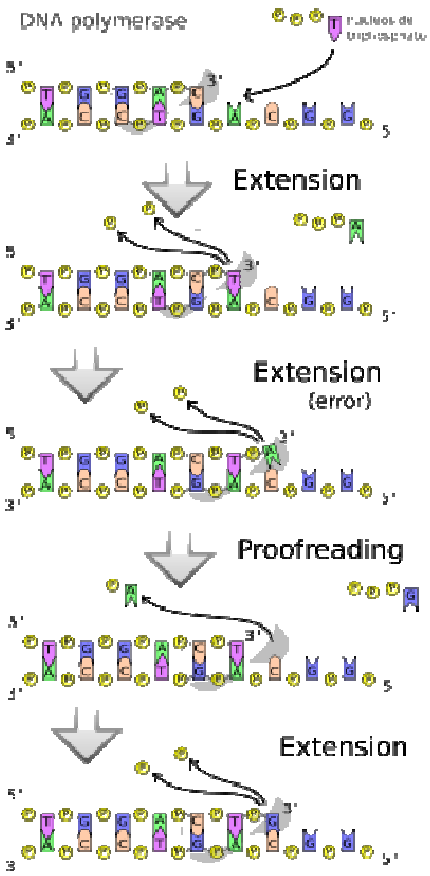


Figure 33: DNA polymerase extension and proofreading during replicative extension. Polymerase effects elongation by sequential addition of dNTPs to 3' sticky ends, erroneous dNTP additions are excised by polymerase exonuclease activity and extension continued as normal [191].

The choice of polymerase can greatly impact resultant base substitution rates; with errors of 10^{-2}- >= 10^{-6} and 10^{-6}-10^{-7} reported for non-proofreading and proofreading facilitated polymerases respectively. An array of chimeric polymerases have been developed for this application including *Pfu*, *Vent* , *Deep Vent* and *UlTm,* however *Pfu* polymerase has proved to be particularly effective displaying a ~10-fold increase in fidelity compared to non-proofreading *Taqs* [193]. It is estimated that the typical polymerase error rate is $1x10^{-4}$; one error may be expected per every 1,000 nucleotides during a twenty cycle amplification. This is a somewhat optimistic figure however, as increases in cycle number greatly increase error rates and the exponential nature of PCR means that mutations incurred early during the reaction process are amplified at each

subsequent cycle. As such final error rates may exceed these estimations greatly [131,194].

The significance of fidelity with regard to PCR amplification depends on the precise application. Indirect characterisation of amplicons by size or even direct characterisation via sequencing and nucleic acid hybridisations are unlikely to be significantly effected by polymerase fidelity due to the relatively low signal strength of aberrant amplicons (assuming reasonable enzyme functionality, cycle numbers and standard amplicon sizes). This is a very important consideration however for applications derived from single molecules or rare targets present in heterogeneous samples [170].

For some applications requiring PCR, circumstances dictate that amplification be performed in the presence of polymerase-inhibitors. A number of polymerases which display increased resistance to common inhibition have been modified using both "directed evolution" and "domain swapping". Ghadessy *et al*., used compartmentalisation of self-replication (directed evolution) to enhance resistance of *Taq* polymerase to heparin inhibitor by 130-fold, for improved amplification of blood samples (see Figure 34), while Pavlov used domain swapping or tagging to combine the protein domains with polymerase catalytic sites to facilitate high-processivity amplification in the presence of high salt concentrations, phenol, blood and intercalating dye inhibitors. These modified polymerases will no doubt become important to the development of PCR, facilitating removal of costly pre-PCR steps and improved real-time PCR amplification efficiency [195,196].

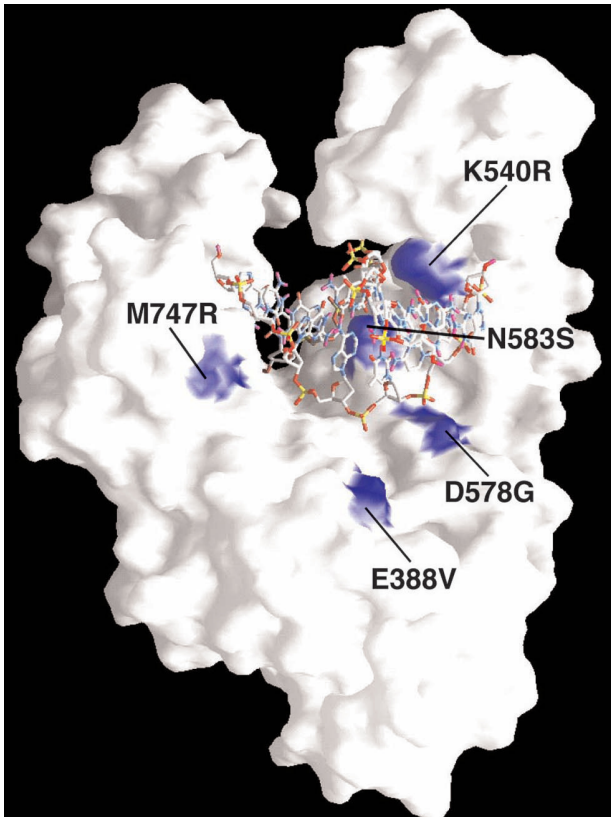


Figure 34: Surface model of the main polymerase domain in *taq* polymerase. Compartmentalized directed evolution was used to identify heparin-resistant inducing mutants (shown in blue)with non-synonymous substitution and positions described [195].

These modified enzymes and novel reaction adjuvants have allowed PCR to be performed with ever-increasing reliability increasing the utility of PCR. A wide range of PCR variants including allele-specific-, nested-, reverse-transcriptase-, rapid-amplification- cDNA-ends-, methylation-specific-, direct- and asymmetric-PCR have been developed, however real-time PCR and multiplex PCR amplification techniques (used during the course of this work) have especially acted to revolutionise PCR utility by allowing increasing sensitivity detection and reduced costs [197-206]. The following section gives an overview of real-time and multiplex PCR techniques, including advantages and limitations of these applications.

4.1.2.1 Real-time PCR (RT-PCR)

PCR as previously described, assess yields following completion of the three PCR reaction phases (exponential, linear and plateau) and is referred to as an 'end-point' method. Due to renaturation competition between complementary product strands, the linear phase introduces considerable variability even between replicates. Post-PCR processing also introduces further pipettor error and detection methods used (i.e. agarose gel / ethidium bromide staining) are also often less than sensitive, varying by as much as 10-fold for band densitometry based measures. As such, end-point procedures are ill-suited to PCR analyses that require high accuracy and precision. Real-time PCR, also termed quantitative PCR (qPCR), was developed in 1996 to address this weakness and improve data yields for PCR amplification experiments [207,208].

Real-time PCR uses fluorescent dyes or probes in association with instrumentation equipped with dedicated fluorescence to monitor the accumulation of amplicons at each PCR cycle [209]. These procedures follow the same three-phase pattern as end-point PCR, however during the annealing / extension phase, either intercalating dye or fluorescent target-specific probes may be used for high-sensitivity target detection [208]. Fluorescence data pertaining to the most stable exponential phase is then used to calculate amplicon yields and determine reaction efficiencies [207].

Aside from the higher precision capacity afforded by exponential data acquisition, the amplicon detection methods employed in real-time PCR also effect sensitivity and precision. A whole host of probe types including hybrid, hydrolysis, taqman, scorpion probes and molecular beacons are commonly used for RT-PCR determination. These exploit fluorescence resonance energy transfer (FRET) and fluorescent quenching for

signal detection during annealing and / or extension. Target-specific probe implementation offers an advantage in terms of specificity over intercalating dyes, however probe synthesis is expensive and a number of probe-induced complications may also be incurred. Apart from increased optimisation requirements (concentration adjustment, hydrolysis and quenching efficiency determination), the introduction of additional probe moieties increases molecular crowding and reduces the capacity for multiplexing. Additionally some sequences may also simply be unsuited to internal probe hybridisation due to GC distribution or secondary structure constraints [208] , as such intercalating dyes such as SYBR Green I, SYTO-13, SYTO-82, SYTOX Orange, TO-PRO-3, -3, POPO-3 and BOBO-3 have been developed for amplicon detection and quantitation [127,210].

Like ethidium bromide, SYBR Green 1 (the most commonly used intercalating dye) displays preferential binding with double stranded DNA. It also displays a much higher binding affinity (100-fold) and fluorescence signal (1000-fold) however, making it more suitable for high-sensitivity detection [208]. The degenerate binding mechanism of intercalating dyes offers huge benefits in terms of reduced RT-PCR costs and the ease of experimental design and optimisation, although it also results in lower specificity relative to probe detection [208].

Newer generation dyes such as SYTO-13 and SYTO-82 offer improved detection sensitivity and alleviate some of the problems previously associated with use of intercalating dyes including preferential binding with GC-rich targets and intercalation-induced PCR inhibition [210]. After real-time PCR amplification, instruments may be programmed to perform melt-peak / curve analysis. Using this approach dsDNA

amplicons are melted by incremental temperature increases resulting in fluorescence reduction with escalating denaturation and dissociation of intercalating dyes (or probes). By monitoring the associated change in fluorescence the point at which the melting temperature for each amplicon is reached may be identified, thereby facilitating target identification even within multiplex formats [211]. Additionally the development of high-resolution, melt-curves (facilitated by an increased rate of fluorescence detection) allows high-precision curves to be derived which may then be compared to known standards for SNP genotyping determination [212].

4.1.2.2 Multiplex PCR

Multiplex PCR facilitates simultaneous amplification of multiple targets within a single reaction format. This higher-dimensionality structure affords a number of benefits which have driven wide-spread adoption of this technique. The largest benefit may be seen in terms of cost; which applies to consumable use, preparation time and maximum utilisation of limited target templates. As multiplex amplification is more demanding than singleplex amplification (covering a larger expanse of genomic / cDNA targets) it also offers an improved capacity for target quality determination. Additionally multiplex amplification facilitates the inclusion of internal reaction controls that act to differentiate between "complete" PCR and "target-specific" PCR failure leading to more robust sample analysis [9]. Multiplex amplification incurs a number of drawbacks which may, however, impede or prohibit effective target specific amplification. As discussed in chapter 3, primer design is significantly more complex for multiplex amplification; therefore if amplification is to be successful, a number of constraints outside the general remit of singleplex primer design must be adhered to. The propensity toward hetero-

dimer formation for all primers in the set must be assessed and minimised to a manageable level, inter-primer T_M range should be relatively narrow and amplicon lengths must also be divergent enough to facilitate differentiation between amplicons but narrow enough to retain thermocycle profile compatibility [162]. While consideration of such parameters is not particularly complex, sequence and primer design constraints mean universally applicable T_Ms, minimal dimerisation profile and amplicon production in the desired size range cannot always be achieved. The complexity of the problem also increases dramatically with multiplex dimensionality, making larger multiplex construction difficult.

Amplification robustness decreases with increasing multiplex dimensionality however and as such a wide range of singleplex assays are recommended to identify potential primer pair weaknesses and failures prior to multiplex inclusion. T_A profile analysis may be used to determine optimal experimental primer pair T_A, while efficiency and dynamic range may be used to determine relative yields and reaction sensitivity across a range of target concentrations. This type of dynamic range-efficiency measure is generally applied for expression sensitivity determination; however as primer pair performance suffers upon multiplex inclusion, dynamic range-efficiency assessment can be used to determine whether primer pairs are robust enough to facilitate multiplex inclusion. According to Edwards *et al*., multiplex reactions should be constructed by sequential addition of primer pairs until the full experimental capacity of the multiplex reaction is achieved, however reaction components ($MgCl_2$, dNTP concentration, polymerase) and amplification profiles are subject to the extent of multiplex dimensionality, primer concentrations used and targets generated; as such, optimal conditions will deviate for alternate multiplex

constructs and optimisation procedures performed in this way can be long and arduous [9,213].

The following methods were performed to address the thesis objective "*To perform extensive PCR optimisation for the construction of a number of robust, well characterised, high dimensionality multiplex PCR sets*". During the course of this work all primer pairs were evaluated extensively using singleplex end-point and real-time PCR to determine amplification characteristics and identify any weaknesses in terms of specificity or efficiency. All high-performing primer pairs were then constructed into the highest dimensionality multiplex formats appropriate for discrimination using available instrumentation. Multiplex reactions were subsequently optimised to create relatively equimolar amplification profiles suitable for downstream SAT. The following section describes the experimental protocols implemented during the course of this work.

4.2 Methods

All methods were performed as described and reagents purchased from Qiagen (Qiagen Ltd, West Sussex, UK) unless otherwise stated.

4.2.1 Cell culture

OE21 human Caucasian oesophageal squamous cell carcinoma cell lines were cultured as described (see appendix 1.2.1 Cell Culture). DNA extraction was performed according to the manufacturer's instructions using the Qiagen QIAamp DNA Mini Kit and quantification / qualification performed using the Warburg and Christian spectrophotometric molar extinction coefficient absorption method (Eppendorf BioPhotometer) Samples displaying purity scores of >2 and 1.7-1.9 for A_{260}/A_{228} and A_{260}/A_{280} measurements were used for subsequent PCR amplification experiments.

4.2.2 Primer Handling and Processing

IGF multiplex set primer sequences described in chapter 3 containing 5' biotin reverse primer modifications were synthesized by Thermo-Electron (OD of 0.2µM, RP-HPLC purification). Lyophilised primers were rehydrated to an approximate concentration of 200µM using sterile nuclease-free Tris-EDTA (TE) buffer pH7.8 in accordance with datasheet instructions and handling specifications. Subsequent spectroscopic measurement and TE volume adjustment was used to normalise primers to a final concentration of 100µM (Eppendorf BioPhotometer). All spectroscopic measurements were performed in triplicate with OD readings ranging between 0.1-1.0 in accordance

with good standard practice. Once normalised, primer solutions were divided into aliquots of 20µl and stored at -20°C to minimise the probability of cross-contamination.

4.2.3 Single-plex End-Point PCR

In order to assess basic primer-pair functionality; singleplex end-point PCR was performed for all 27 primer pairs using the Qiagen Multiplex PCR and GeneAmp AmpliTaq Gold kits as described Table 17 and Table 18 respectively. GeneAmp Ampli*Taq* Gold kit was purchased from Applied Biosystems (Applied Biosystems Inc**,** Foster City, CA). All amplifications were performed using a Techne TC-512 thermal cycler (Techne Inc, Staffordshire, United Kingdom) and thermal cycle profile A (Table 16). Amplification products were separated by agarose gel electrophoresis and stained using ethidium bromide (electronic appendix 1.2.3 Agarose Gel Electrophoresis). Syngene Gene Genius Bioimaging System in association with GenSnap and GeneTools software (Synoptics Ltd, Cambridge, UK) were used to visualise and analyse products by molecular weight marker comparative analysis.

Table 16: Thermal Cycle Profile A

Step	Temp (°C)	Duration (min/sec)	Cycle no.
Preheated Lid	105	3m00s	1
Initial Denaturation	95	15m00s	1
Denaturation	95	00m30s	30
Annealing	65	00m30s	
Extension	72	00m30s	
Final Extension	68	15m00s	1

Table 17: Qiagen Multiplex PCR Amplification Reaction Mix. Reaction mix "D"

	Component	Volume (µl)
1	Qiagen Multiplex Master Mix	7.5
2	Primer mix (10uM)	2.0
4	DNA Template (190ng/µl)	1.0
5	Deionised nuclease free H_2O	4.5
	Total Volume (µl)	**15**

Table 18: Ampli*Taq* Gold PCR Amplification. Reaction mix "B

	Component	Volume (µl)
1	Deionised nuclease free H_2O	9.90
2	10x Gold buffer	1.50
3	dNTP (200 µM)	0.15
4	Forward Primer (10µM)	1.00
5	Reverse Primer (10µM)	1.00
6	Target DNA (190ng/µl)	1.00
7	Ampli*Taq* gold DNA Polymerase (5 units/µl)	0.45
	Final Volume (µl)	**15.00**

4.2.4 Repeat Region Amplification Optimisation

Further optimisations were performed for primer pairs (IGF1_9, IGF1_10 and BP3_14) displaying non-specific amplification using six alternate amplification protocols termed A, B, C, D, E and F. These are described as follows:

- **A = Standard *taq,* $MgCl_2$ restricted, manual hot start**

Protocol A was composed of reactants as detailed (Table 19). $MgCl_2$ was excluded from the reaction mix in an attempt to minimize $MgCl_2$ facilitated secondary structure stabilisation and a manual hot start implemented. This was performed by withholding polymerase addition until reactants had reached 95°C (during the initial thermal cycle denaturation step) and should act to reduce primer-dimer and or non-specific hybridisation and amplification.

Table 19: Reaction mix A - Standard *taq,* $MgCl_2$ restricted, manual hot start

	Component	Volume (µl)
1	Deionised nuclease free H_2O	11.5
2	buffer	1.50
3	dNTP (200 µM)	0.15
4	Forward Primer (10µM)	0.2
5	Reverse Primer (10µM)	0.2
6	Target DNA (190ng/µl)	1.00
7	*Taq* DNA Polymerase (5 units/µl)	0.45
	Final Volume (µl)	**15.00**

- **B = Time-release Ampli*Taq* -gold, $MgCl_2$ restriction**

Protocol B reactants were composed as detailed (Table 18). Again $MgCl_2$ was excluded from the reaction mix in an attempt to minimize $MgCl_2$ facilitated secondary structure stabilisation. A higher-processivity, higher-fidelity Ampli*Taq* polymerase was also used to try to reduce polymerase mediated nucleotide inclusions or exclusions (indel). In addition a time-release thermal cycle profile which implements a short 5 minute initial denaturation and increased cycle number (35cycle) profile was introduced (Table 20). This allowed slow activation of chemically inactivated Ampli*Taq* gold polymerase over a

number of cycles to allow polymerase concentrations to mirror those of target production and should reduced excess-polymerase induced non-specific amplification.

Table 20: Time-release Thermal Cycle Profile B

Step	Temp (°C)	Duration (min/sec)	Cycle no.
Preheated Lid	105	3m00s	1
Initial Denaturation	95	5m00s	1
Denaturation	95	00m30s	35
Annealing	65	00m30s	
Extension	72	00m30s	
Final Extension	68	15m00s	1

- **C and F = Time-release Ampli*Taq*-gold, $MgCl_2$ restricted with 5% and 10% DMSO respectively.**

Again this protocol used Ampli*Taq* gold and excluded $MgCl_2$ however destabilising agent DMSO (Sigma Aldrich Dimethyl sulfoxide biotech. grade, 99.8%) was also added to reduce hydrophobic forces and the stability of target secondary structures which may contribute to non-specific amplification by deletion mutagenesis mechanisms [214]. Ampli*Taq* Gold Reaction mix "B" (Table 18) was used with addition of 5% and 10% DMSO and H_2O adjustment for protocols C and F respectively. Thermal cycle A was again used (Table 16) to facilitate amplification.

- **D and E = Qiagen multiplex master mix, 3mM $MgCl_2$, MP-factor (with 5% Q solution for reaction mix E)**

Reaction mix "D" was used for amplification of targets as described (table 17). The Qiagen multiplex PCR kit is supplied with a premixed master, containing HotStar

polymerase, synthetic factor MP and $MgCl_2$. Factor MP increases the local concentration of primers at the DNA template and should help stabilise specifically bound primers. Reaction mix "E" also contains 5% Q-solution isostabilizing agent which can act to improve amplification of GC rich targets and those containing a high degree of secondary structure [215]. 5% Q-solution was added to these reactions and H_2O adjusted accordingly. Thermal cycle profile A (Table 16) was used for amplification in both instances.

4.2.5 Single-plex Real-Time PCR

The Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories Ltd., Hertfordshire, UK) was used to perform high-sensitivity singleplex real-time PCR amplification of all primer pairs. Multiplex critical parameters; annealing temperature and amplification efficiency / dynamic range were investigated using this highly sensitive technique. Note: A range of optimisations including template (0.025-250ng/µl) and primer (50-900nM) concentrations were first performed for identification of optimal conditions prior to annealing temperature profile or standard curve determination. Primer and template negatives were also prepared and included in each analysis plate.

4.2.6 Annealing Temperature Profile Determination

In association with OE21 target and single primer pairs; Bio-Rad's IQ SYBR Green Supermix (with high-fidelity antibody-mediated hot start polymerase and high-sensitivity proprietary intercalating dye) was used for amplification and detection of PCR products. Master mixes were constructed for each primer pair to minimize intra-sample variations which may affect yields in a manner independent of annealing temperature effects. A gradient thermal cycle profile facilitating amplification with annealing temperatures

spanning an 8.9°C range was constructed as described (Table 22 and Table 23). For each prime pair, 10ul of master mix was added to each well (B through F) which corresponded to T_As of 59.9, 62.2, 65.1, 67.5 and 68.8°C respectively. Amplification curves, melt peaks and electrophoretic analysis was used to assess specificity and kinetic conformance of the reaction. This information in association with relative yields across the T_A range were used to determine optimal experimental T_A for each primer pair.

Table 21: Real-time PCR IQ SYBR Green Reaction mix.

	Components	**1X Volume (µl)**	**5.2X Volume (µl)**
1	2X IQ SYBR Green Supermix	5	26
2	Primer mix (2uM forward and reverse primer)	1	5.2
3	Template (conc. 13.62ng/µl)	1.834	9.53
4	Deionised nuclease free H_2O	2.2	11.44
	Total Volume (µl)	**10**	**52.0**

Table 22: Real-time PCR Thermal Cycle Profile.

Step	Temp (°C)	Duration (min/sec)	Cycle no
Preheated Lid	105.00	3m00s	1
Initial Denaturation	95.00	10m00s	1
Denaturation	95.00	00m30s	40
Annealing	65.00*	00m30s	
Extension	72.00	01m00s	
+ Plate read			
Denaturation	95.00	01m00s	1
Annealing	40.00	01m00s	1
Melt Curve	From 65.0-95.0 for 0.01s at 0.2°C increment + plate read		

*During real-time T_A optimisation; annealing temperatures as described (Table 23) were used for sample tests of each primer pair. A T_A of 65.00°C was used for amplification of efficiency / dynamic range samples.

Table 23: Real-time PCR Thermal Cycle Annealing Temperature Gradient Profile.

Plate Row Identifier	B	C	D	E	F
T_A Gradient (°C)	68.8	67.5	65.1	62.2	59.9

4.2.7 Efficiency / Dynamic Range Determination

The standard curve method was used for primer-pair efficiency / dynamic range determination. A ten-fold serial dilution of OE21 target DNA (6.16 × 10-1-6.16 × 10-2 fMol) was carried out and an additional two samples containing 3.85 and 2.31fMol of template around the optimal target concentration prepared to create a six-point profile of each primer pair (see Table 25). Master mixes containing ABI Power SYBR Green PCR master mix (with high-fidelity antibody-mediated hot start polymerase and high-sensitivity proprietary intercalating dye) were constructed for each primer pair to minimise intra-sample variations which may affect yields in a manner independent of template concentration (Table 24). Standard Curves were prepared and analysed in duplicate to avoid inter-run and inter-plate variability. Bio-Rad CFX software was used to construct standard curves and carry out efficiency, slope and r^2 determination for each pair. All these features were evaluated to assess primer pair performance and sensitivity. Amplification curves, melt peaks and electrophoretic analysis was used to assess primer pair specificity and kinetic adherence as before.

Table 24: Real-time PCR ABI Power SYBR Green reaction mix.

	Components	1X Volume (µl)	6.3X Volume (µl)
1	2X Power SYBR Green PCR master mix	5	31.5
2	Primer mix (2uM forward and reverse primer)	1	6.3
3	Template (variable conc.)	1.32	8.316
4	Deionised nuclease free H_2O	3.2	20.16
	Total Volume (µl)	**10**	**63**

Table 25: Real-time PCR Standard Curve Template concentrations and dsDNA copy number.

Total Target DNA Concentration (fMol)	dsDNA Copy number
$6.16x10^{1}$	$1.24x10^{5}$
6.16	$1.24x10^{4}$
$6.16x10^{-1}$	$1.24x10^{3}$
$6.16x10^{-2}$	$1.24x10^{2}$
3.85	$7.72x10^{3}$
2.31	$4.63x10^{3}$

4.2.8 Multiplex PCR

Multiplex PCR amplification and analysis was performed using Techne TC-512 PCR system and Bio-Rad's Experion Automated Electrophoresis System (Bio-Rad Laboratories Ltd., Hertfordshire, UK) respectively. Ampli*Taq* gold, $MgCl_2$ restriction reaction mix was used for amplification of the 2plex primer mix while Qiagen's multiplex PCR mix was used for amplification of all other multiplex primer mixes. Extensive optimisation of both reaction mixes and thermal cycle segment times were required for each multiplex amplification. Adjustment was performed by sequential increase of annealing, extension and denaturation times until optimal conditions for each multiplex were reached. Reaction mix adjustment in terms of polymerase selection was instructed from previous singleplex amplification, while a number of other optimisations regarding template concentration and total primer concentration were also made for the largest 14-plex reaction. Final reaction mixes (as listed Table 26, Table 27 and Table 28) and thermal cycle profiles (Table 29, Table 30 and Table 31) were used to produce optimised 2-plex, 5-plex, 6-plex and 14-plex respectively.

Table 26: 2-plex multiplex reaction mix.

GeneAmp Gold PCR Reagent Kit	Volume (µl)
GeneAmp 10X PCR Gold Buffer	3.0
dNTP (200µM)	0.6
$MgCl_2$ (25mM)	2.4
*Primer mix	3.0
Ampli*Taq* Gold DNA Polymerase (5 units/µl)	0.9
Target DNA	0.52 (100ng)
DiNF H_2O	21.08
Total vol	**30**

Table 27: 5-plex, 6-plex and 14-plex multiplex reaction mixes

Qiagen multiplex PCR kit	5-plex (µl)	6-plex (µl)	14-plex (µl)
2x QIAGEN Multiplex PCR Master Mix	12.5	12.5	25
Primer mix*	1.25	1.25	2.5
Q soln (5X)	2.5	2.5	5.0
Target	2.6 (0.5µg)	2.6 (0.5µg)	5.2 (1µg)
DiNF H_2O	4.9	4.9	13.8
Total vol.	**25**	**25**	**50**

Table 28: Primer Pair reaction mixes for final 2-plex, 5-plex, 6-plex and 14-plex PCR amplifications. Final primer pair concentrations ranged from 0.03-0.1, 0.05-0.177, 0.04-0.15µM for 5-plex, 6-plex and 14-plex respectively. An equimolar concentration of 0.1µM per primer was used for the 2-plex amplification.

Multiplex identifier	Primer Pair Identifier	Primer Mix, Primer Pair Conc. (µl)	Final Primer Pair Conc. (µM)	Final total Primer Conc. (µM)
2-plex	IGF1_9	1.0	0.10	0.400
	IGF1_11	1.0	0.10	
5-plex	BP1_3/4	1.0	0.10	0.392
	BP3_10/11	0.3	0.03	
	IGF1_5	1.0	0.10	
	IGF1_2	1.0	0.10	
	BP3_1	0.62	0.062	
6-plex	BP3_13	0.5	0.05	0.341
	BP1_5	2.0	0.20	
	IGF1_5	0.5	0.05	
	BP3_12	1.77	0.177	
	IGF1_7	0.6	0.060	
	IGF1_3	1.45	0.145	
14-plex	IGF1_8	2.5	0.125	2.00
	BP3_9	1.6	0.08	
	BP1_1	1.6	0.08	
	BP3_7	1	0.05	
	BP1_3/4	1.5	0.075	
	BP3_10/11	0.8	0.04	
	BP3_3	1.5	0.075	
	BP1_2	1	0.05	
	BP3_6	1	0.05	
	IGF1_1	1	0.05	
	IGF1_13	1	0.05	
	BP3_4	1.5	0.075	
	IGF1_4	1	0.05	
	IGF1_12	3	0.15	

Table 29: 2-plex Thermal Cycle Profile.

Step	Temp (°C)	Duration (min/sec)	Cycle no.
Preheated Lid	105	3m00s	1
Initial Denaturation	95	5m00s	1
Denaturation	95	1m00s	40
Annealing	65	1m00s	
Extension	72	1m00s	
Final Extension	68	15m00s	1

Table 30: 5-plex and 6-plex Thermal Cycle Profile.

Step	Temp (°C)	Duration (min/sec)	Cycle no.
Preheated Lid	105	3m00s	1
Initial Denaturation	95	15m00s	1
Denaturation	95	50s	35
Annealing	65	2m30s	
Extension	72	1m15s	
Final Extension	68	15m00s	1

Table 31: 14-plex Thermal Cycle Profile.

Step	Temp (°C)	Duration (min/sec)	Cycle no.
Preheated Lid	105	3m00s	1
Initial Denaturation	95	15m00s	1
Denaturation	95	50s	35
Annealing	65	3m00s	
Extension	72	1m30s	
Final Extension	68	15m00s	1

4.2.9 MetaPhor™ Gel Electrophoresis

High resolution MetaPhor agarose gel (Cambrex Corp., Charles City, IA) electrophoresis with ethidium bromide staining was used for initial separation and visualization of smaller dimensionality multiplex amplicons (to up 5-plex). The technique was applied using 5% MetaPhor agarose and 1xTBE (100 Volts for 3hours) in accordance with manufacturers' instructions to achieve a resolution of 16bp in the range of 72-500bp. A number of amendments including use of up to 8% gel in association with 3 × TBE, 24 hours electrophoresis at 4@C and multiple buffer changes were used to increase resolution.

4.2.10 Experion Electrophoresis

Bio-Rad's Experion automated gel electrophoresis system and associated Experion 1K DNA Analysis kit was used to facilitate increased resolution, sensitivity and precision for multiplex amplicon analysis. Analysis was performed in accordance with manufacturers' instructions to achieve 5bp resolution (<160bp). All reagents and chip were subsequently cooled to 0°C to achieve a slightly increased 4bp resolution. Lower temperature electrophoresis has been successfully used by a number of groups including Fanali *et al.,* Tsai *et al.,* and Chen *et al.,* to facilitate increased electrophoretic resolution of this nature [216-218]. 1µl of undiluted PCR product was analysed per well for each multiplex. Amplicon yields were determined by automated comparison between amplicon peaks and Experion DNA 1K ladder using Experion software.

The following calculation is used to determine the relative yield produced for each prime pair within optimised multiplex formats, with adjustment for primer pair concentration in each instance. For each multiplex constituent primer pair:

- Primer Pair Yield / Primer Pair concentration = X
- X/ Largest X in multiplex = Relative end-point efficiency

The following calculation was performed to determine the largest % difference in yield for all primer pairs amplified within individual multiplexes.

- Lowest primer pair yield / (Highest primer pair yield/ 100) = Y
- 100 -Y = % divergence between largest and smallest yielding primer pairs

4.3 Results

4.3.1 Single-plex PCR amplification

All primer pairs were amplified singly using uniform PCR reactant mixes and thermal-cycling conditions. Specificity and yield of individual primer pairs under set conditions were examined to identify potentially problematic pairs which may require specialised singleplex amplification or adjustment prior to multiplex inclusion. The following ethidium bromide stained gels (Figure 35, Figure 36 and Figure 37) display amplicons procured using high fidelity, high processivity Ampli*Taq* gold polymerase as previously specified (Table 16 and Table 18). Singleplex amplicons derived using the Qiagen Multiplex PCR kit are displayed in electronic appendix Figure 1.

25 of the 27 amplicons displayed high-specificity amplification, producing targets of the anticipated size (as predicted by UCSC *in silico* PCR analysis). Repeat region targets IGF1_10 and BP3_14 displayed non-specific amplification and multiple bands, however IGF1_9 (also a repeat region target) produced a single strong amplicon band. Primer pair BP3_1 produced relatively weak amplicon yields relative to other primer pairs.

Ladder IGF1_1 IGF1_2 IGF1_3 IGF1_4 IGF1_5 IGF1_6 IGF1_7 IGF1_8 IGF1_9 IGF1_10

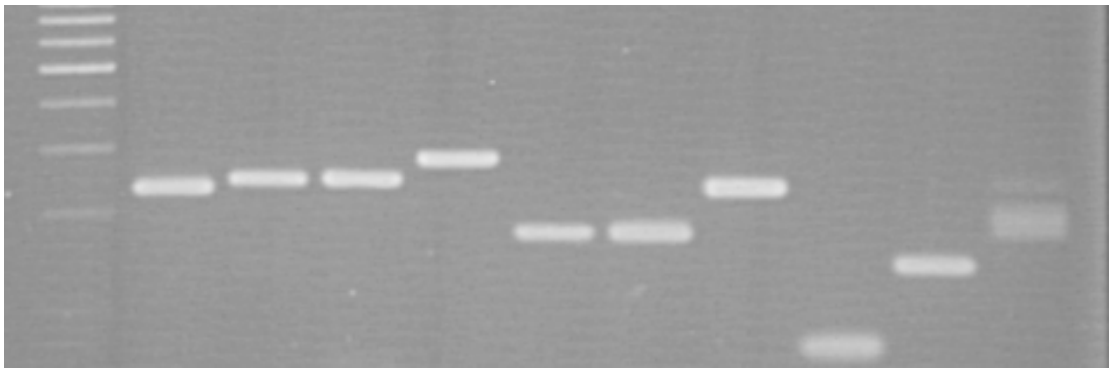


Figure 35: Singleplex PCR amplification of target loci. Ampli*Taq* Gold mediated PCR amplification was performed under uniform conditions for individual primer pairs and resultant product separated and visualised using 3% agarose gel electrophoresis and ethidium bromide staining. Primer pair identifiers are listed above associated lanes.

Ladder IGF1_11 IGF1_12 IGF1_13 BP1_1 BP1_2 BP1_3/4 BP1_5 BP3_1 BP3_3 BP3_4

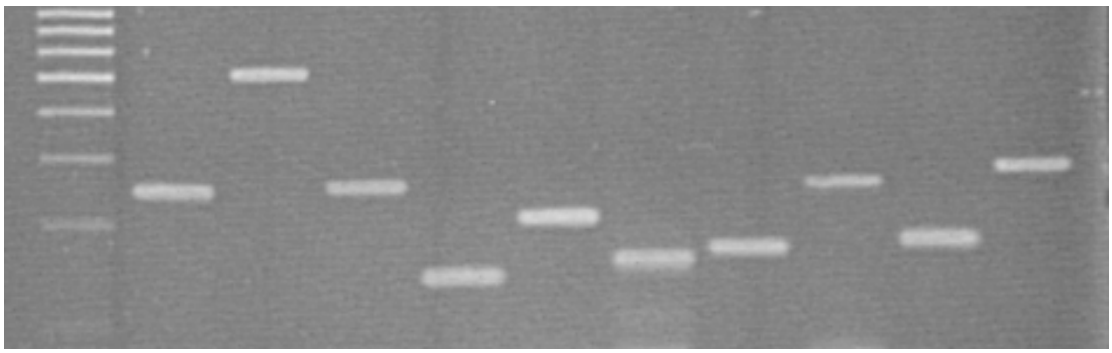


Figure 36: Singleplex PCR amplification of target loci. Ampli*Taq* Gold mediated PCR amplification was performed under uniform conditions for individual primer pairs and resultant product separated and visualised using 3% agarose gel electrophoresis and ethidium bromide staining. Primer pair identifiers are listed above associated lanes.

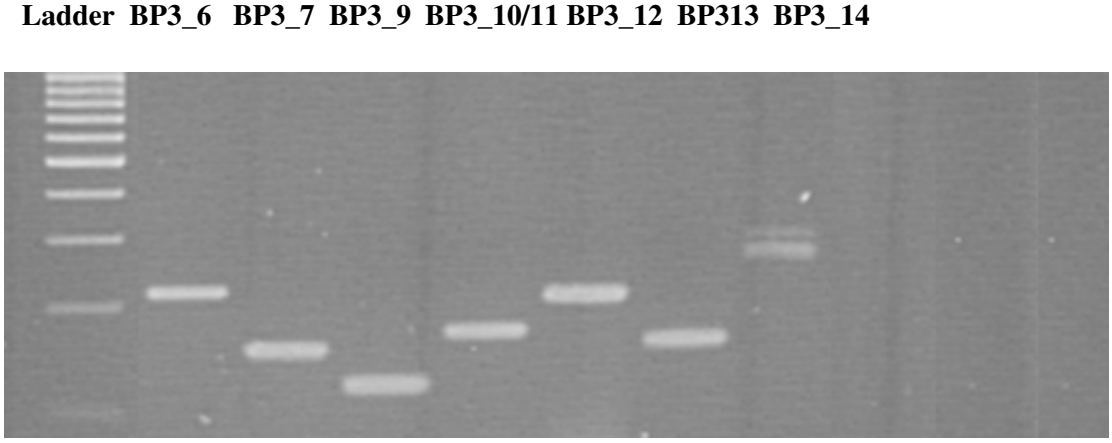


Figure 37: Singleplex PCR amplification of target loci. Ampli*Taq* Gold mediated PCR amplification was performed under uniform conditions for individual primer pairs and resultant product separated and visualised using 3% agarose gel electrophoresis and ethidium bromide staining. Primer pair identifiers are listed above associated lanes.

4.3.2 Problematic Primer Pair Optimisation

Primer pairs IGF1_9, IGF1_10 and BP3_14 with targets spanning putative repeat regions were optimised to reduce aberrant band formation. A number of alternate strategies including use of alternate enzymes (standard *taq*, Ampli*Taq* Gold and HotStar DNA polymerases), addition of adjuvants (DMSO and Q-solution) and magnesium chloride restriction were implemented to reduce erroneous amplification and minimise secondary structure formation. The following ethidium bromide stained agarose gels (Figure 38, Figure 39, Figure 40 and Figure 41) display optimisation of each primer pair using the six treatments described (see 4.2.4 Repeat Region Amplification Optimisation).

Use of Ampli*Taq* Gold polymerase (treatments B, C and F) facilitated target specific amplification in all instances for IGF1_9. Use of treatment F (Ampli*Taq* gold with 10% DMSO and time-release thermal cycle protocol) significantly increased relative

concentration of the IGF1_10 target amplicon relative to other non-specific products; however total yield was very low. Target specific amplification was not achieved for BP3_14 using the amendments detailed here; additionally use of even 5% DMSO significantly reduced yields.

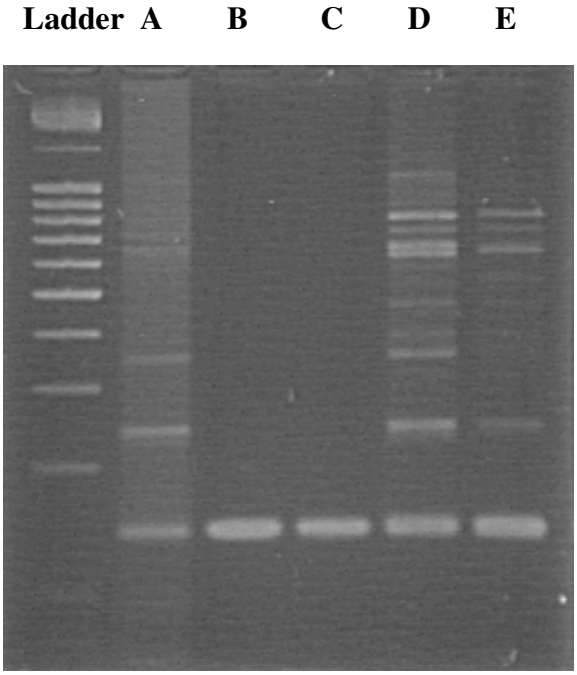


Figure 38 IGF1_9 reaction mix optimisation for amplification of target repeat regions. Treatment A = standard *taq* DNA polymerase, B = Ampli*Taq* Gold polymerase, C = Ampli*Taq* gold polymerase with 5% DMSO, D = Hot star polymerase and MP factor, E = Hot star polymerase, MP factor and Q solution.

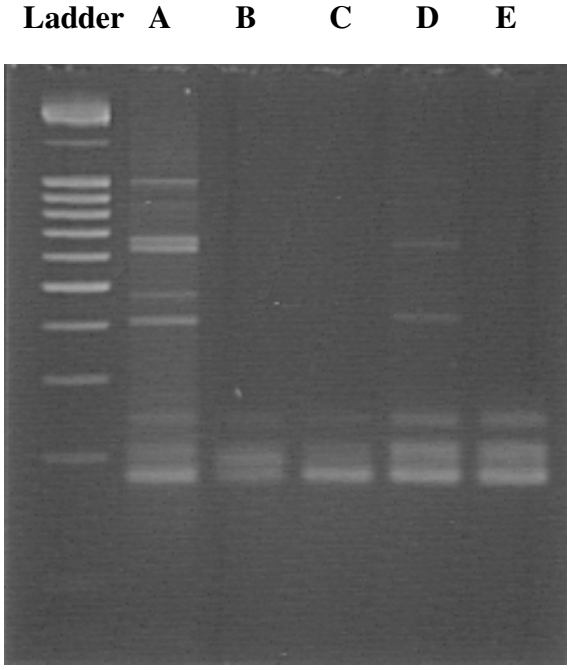


Figure 39 IGF1_10 reaction mix optimisation for amplification of target repeat regions. Treatment A = standard *taq* DNA polymerase, B = Time-release Ampli*Taq* Gold polymerase, C = Time-release Ampli*Taq* gold polymerase with 5% DMSO, D = Hot star polymerase and MP factor, E = Hot star polymerase, MP factor and Q solution.

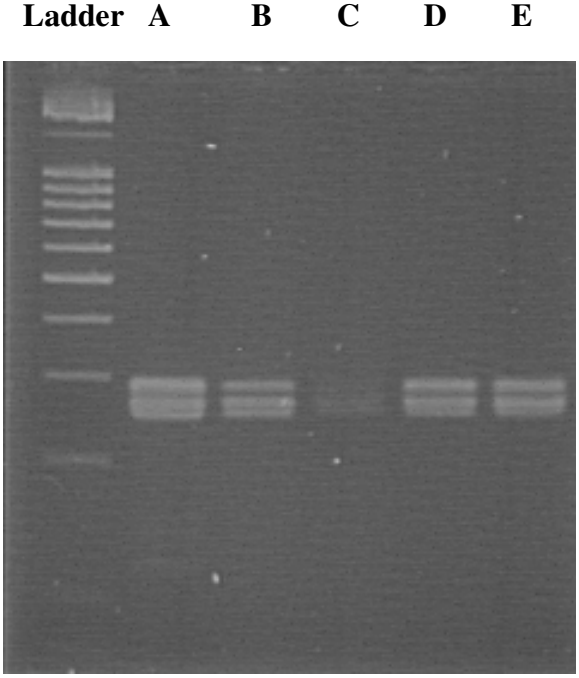


Figure 40 BP3_14 reaction mix optimisation for amplification of target repeat regions. Treatment A = standard *taq* DNA polymerase, B = Time-release Ampli*Taq* Gold polymerase, C = Time-release Ampli*Taq* gold polymerase with 5% DMSO, D = Hot star polymerase and MP factor, E = Hot star polymerase, MP factor and Q solution.

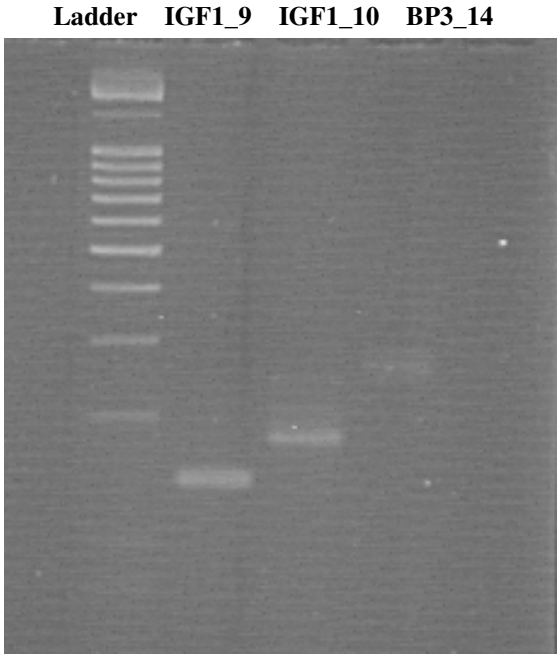


Figure 41 Time-release Ampli*Taq* gold facilitated PCR amplification of repeat region targets for primer pairs IGF1_9 IGF1_10 BP3_14. 10% DMSO (treatment F) was used to reduce secondary structure formation which can result in multiple band formation.

4.3.3 RT-PCR Annealing Temperature Optimisation

All primer pairs were amplified singly using uniform RT-PCR reactant mixes spanning a range of annealing temperatures (59.9-68.8°C) including the putative optimal annealing temperature (~65°C) as described (4.2.6 Annealing Temperature Profile Determination). Amplification curves, melt peaks and electrophoretic analysis was performed to determine whether primer dimer formation and / or non-specific amplification were likely to have had occurred. Optimal annealing temperatures for each primer pair were attributed to those T_As which produced highest relative target yields (assuming adequate amplification specificity at given T_As). RT-PCR amplification curves, melt peaks and annealing temperature profiles for IGF1_1, BP3_1 and IGF1_9 are displayed in full (Figure 42, Figure 43 and Figure 44 respectively).

IGF1_1 amplification curves follow the desired sigmoid shape and melt peaks do not display significant primer dimer or multiple band formation. Largest yield and therefore optimal annealing temperature for IGF1_1 occurs at 65.1°C. IGF1_1 appears to be quite robust with strong amplification across all T_As (59.9-68.8°C) tested (Figure 42). BP3_1 also displayed good amplification however melt peaks displayed an increase in primer dimer formation at the lower 59.9 °C temperature, a feature which was taken into account when assessing yields and assigning optimal T_A (Figure 43).Melt peak profiles clearly display non-specific amplification for primer pair IGF1_9, with alternate products favoured at opposing ends of the T_A spectrum tested. Although the IGF1_9 annealing temperature profile is displayed, results are not indicative of high-specificity target yields and were not used to determine optimal T_A in this case (Figure 44).

Figure 45 displays optimal experimental annealing temperature for high-specificity primer pairs. Primer3 T_A (Panjkovich) consensus method displayed good predictive accuracy with >70% of primer pairs tested displaying optimal performance at the anticipated T_A ~65°C. A slight trend toward underestimation of T_A may be noted with 25% of primer pairs displaying superior performance at 67.5°C. Just one primer pair BP3_10/11 displayed optimal performance outside these temperatures at 62.2°C. Annealing temperature profiles for all other primer pairs are displayed in electronic appendix Figure 5-31.

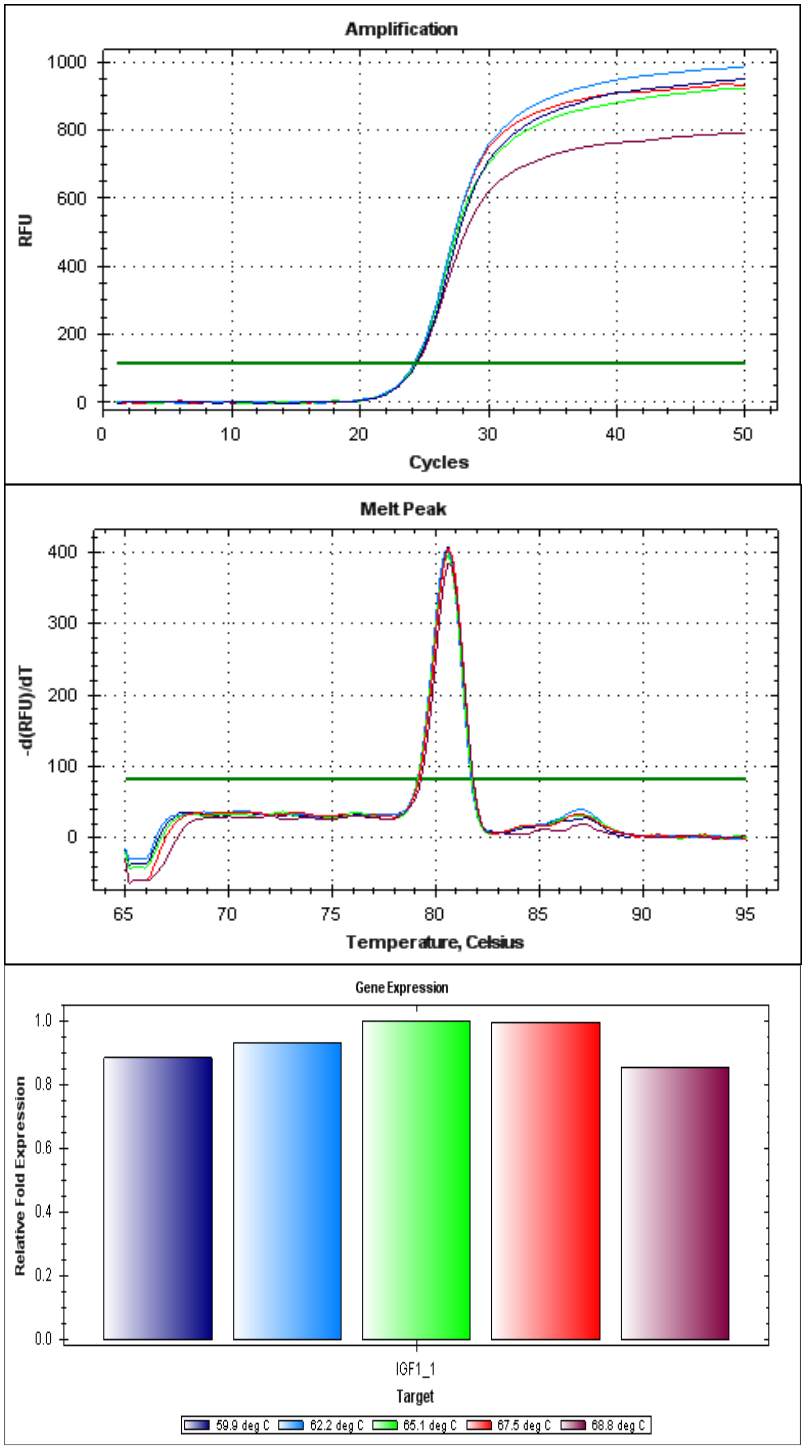


Figure 42: IGF1_1 Amplification Curve, Melt Peak and Annealing Temperature Profiles are displayed. Amplification curves follow the desired sigmoid shape and melt peaks do not display significant primer dimer or multiple band formation. Largest yield and therefore optimal annealing temperature for IGF1_1 occurs at 65.1°C. IGF1_1 appears to be quite robust with strong amplification across all T_As (59.9-68.8°C) tested.

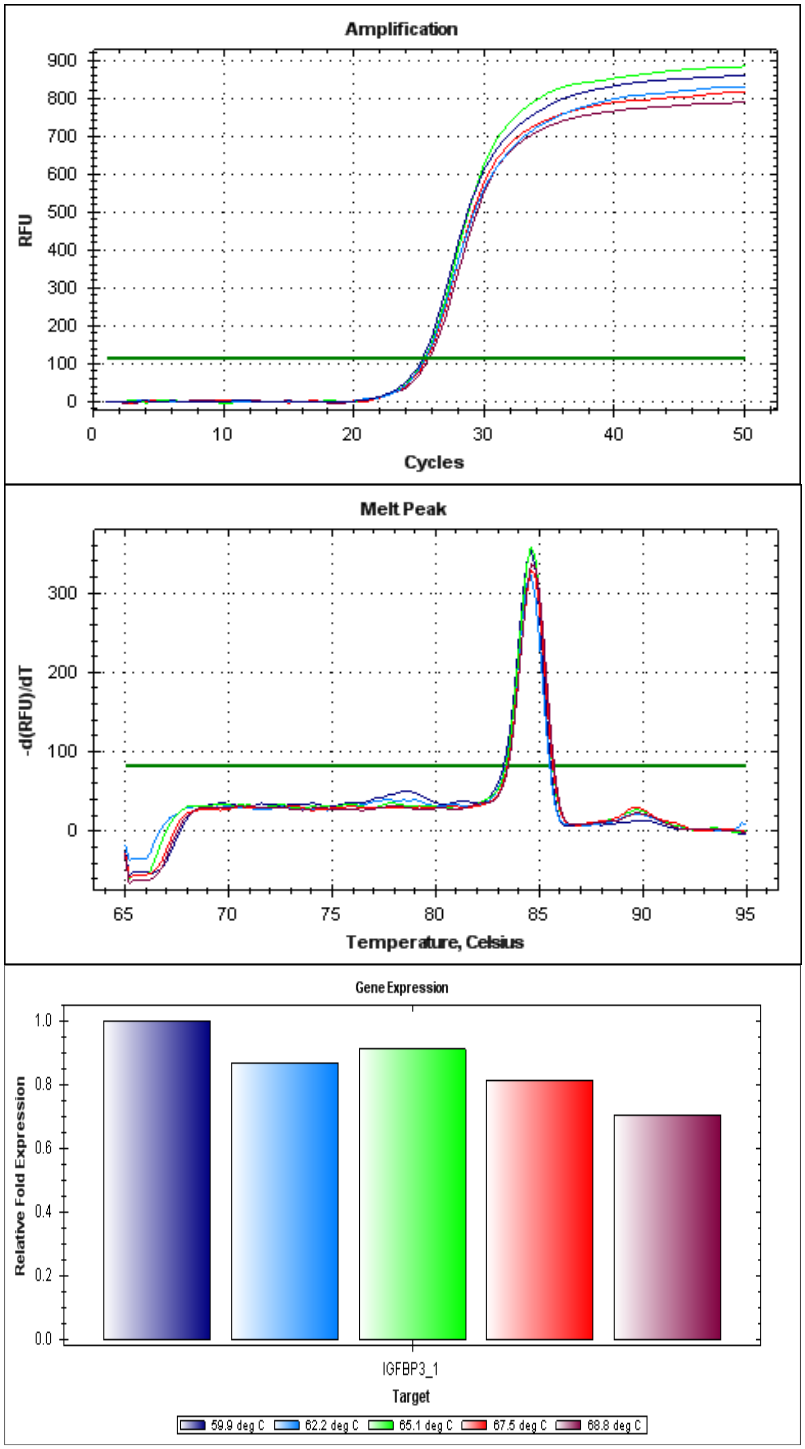


Figure 43: BP3_1 Amplification Curve, Melt Peak and Annealing Temperature Profiles are displayed. Amplification curves follow the desired sigmoid shape however melt peaks do display increased primer dimer formation at the lower 59.9 °C temperature. This was taken into account when assessing yields and assigning optimal T_A.

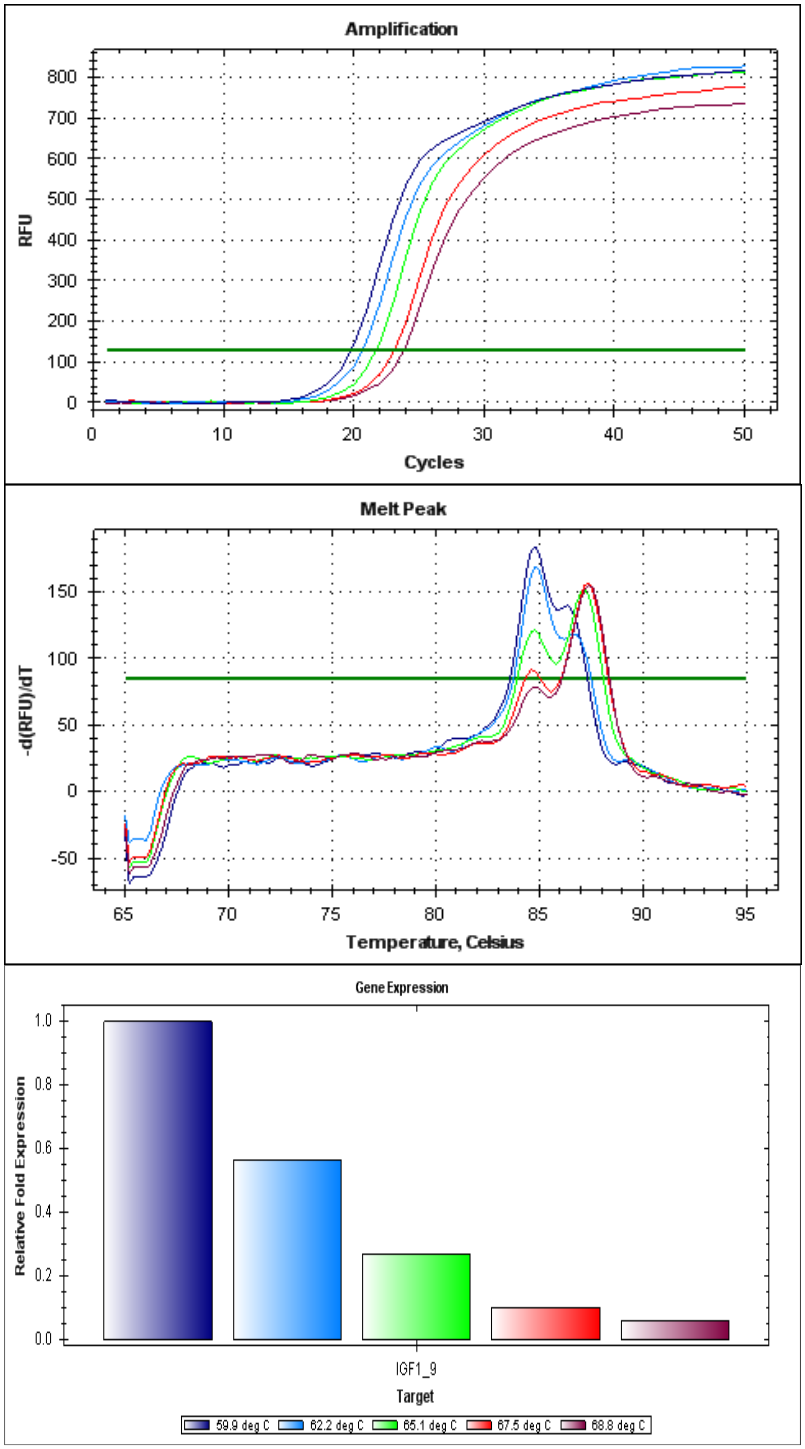


Figure 44: IGF1_9 Amplification Curve, Melt Peak and Annealing Temperature Profiles are displayed. An amplification curve correlation between increasing T_A and C(t) may be observed; a feature unassociated with robust target-specific primer pair functionality. Non-specific amplification is clearly visible by melt peak analysis, with alternate products favoured at opposing ends of the T_A spectrum. Although the IGF1_9 annealing temperature profile is displayed, results are not indicative of high-specificity target yields and therefore may not be used to determine optimal T_A.

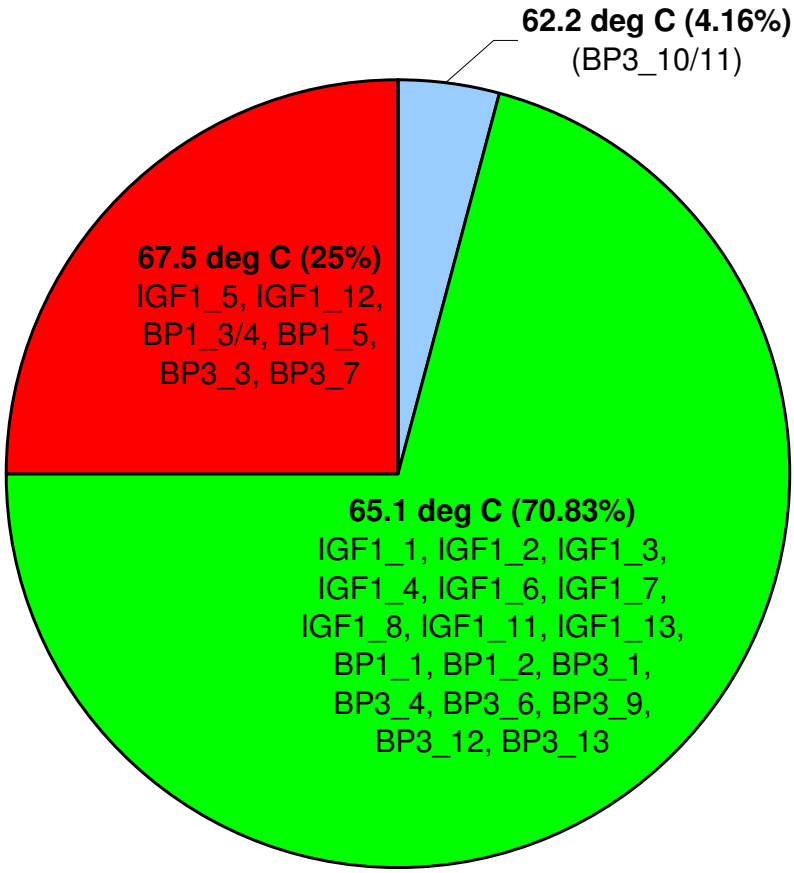


Figure 45: Optimal experimental annealing temperature for high-specificity primer pairs. Primer pair identifiers and the percentage of high-specificity pairs displaying optimal performance at each T_A are shown. >70% of primer pairs displayed optimal amplification at 65.1°C, 25% displayed optimal performance at 67.5°C and just one primer pair displayed optimal amplification performance at 62.2°C.

4.3.4 RT-PCR Efficiency and Dynamic Range

All primer pairs were amplified singly using uniform RT-PCR reactant mixes and thermal-cycling conditions. Efficiency and of individual primer pairs under optimal conditions were examined across a dynamic range four orders of magnitude to identify potentially problematic pairs which may suffer upon multiplex inclusion. A four-point ten-fold serial dilution of OE21 target DNA Amplification (6.16×10^{-6}-16×10^{-3} fMoles) was carried out and an additional two samples containing 3.85 and 2.31fMoles of template around the optimal target concentration prepared to create a six point profile of each primer pair and efficiency, slope and r^2 noted. Visual inspection of amplification curves and melt peak shapes and electrophoresis was used to assess whether primer dimer formation and / or non-specific amplification had occurred. Values within the following ranges; Efficiency = 90-110%, Slope = -3.6 and -3.1 and r^2 = >0.95 are indicative of robust "good" primer pair functionality, primer pairs with scores exceeding set limits in any of these categories was said to display "poor" functionality.

RT-PCR amplification curves, melt peaks and standard curves for high-functionality IGF1_1 and sub-optimal IGF1_8 and IGF1_9 primer pair are displayed in full (Figure 46, Figure 47 and Figure 48). IGF1_1 amplification curves follow the desired sigmoid shape and pattern (approx 3.3 cycles separating 10-fold serial dilutions) and no significant mispriming was visible by melt curve analysis (Figure 46). This primer pair like most others in the IGF multiplex set displayed excellent efficiency over the dynamic range tested with efficiency of 99.3% (slope = -3.338, r^2 = 0.986). Although amplification was target-specific; IGF1_8 demonstrated poor efficiency over the dynamic range tested with efficiency of 114.3% (slope = -3.021 and r^2 = 0.989) (Figure 47).

Amplification curves displayed a sigmoid curve shape but do not follow the anticipated pattern of 3.3 cycles between 10-fold serial dilutions for higher $6.16x10^{1}$fM target samples. IGF1_9 demonstrate poor efficiency of 152.9% (slope = -2.471 and r^2 = 0.775) and specificity over the dynamic range tested (Figure 48). Melt peaks were indicative of poor specificity which was confirmed by agarose gel electrophoresis. Figure 49 displays efficiencies for all primer pairs while Table 32 gives details of efficiency, slope r^2 and functionality designation. Standard curves for all other primer pairs are displayed in electronic appendix Figure 32 to Figure 40.

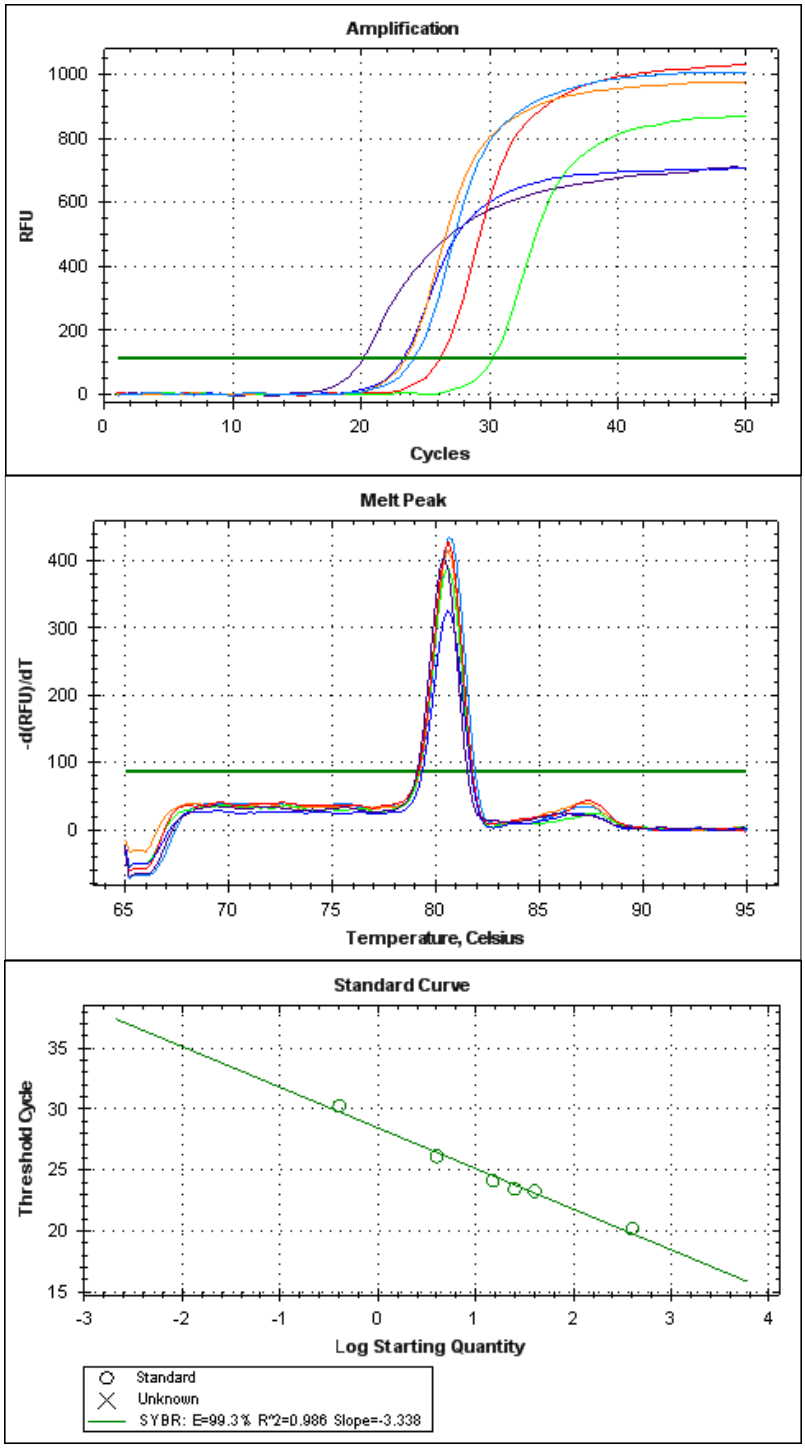


Figure 46: IGF1_1 Amplification Curve, Melt Peak and Standard Curve. Amplification curves follow the desired sigmoid shape and pattern (approx 3.3 cycles separating 10-fold serial dilutions) no significant mispriming is evident by melt curve analysis. IGF1_1 displays excellent efficiency over this dynamic range ($6.16x10^{1}$-$6.16x10^{-2}$fM) with E = 99.3%, slope = -3.338 and r^2 = 0.986.

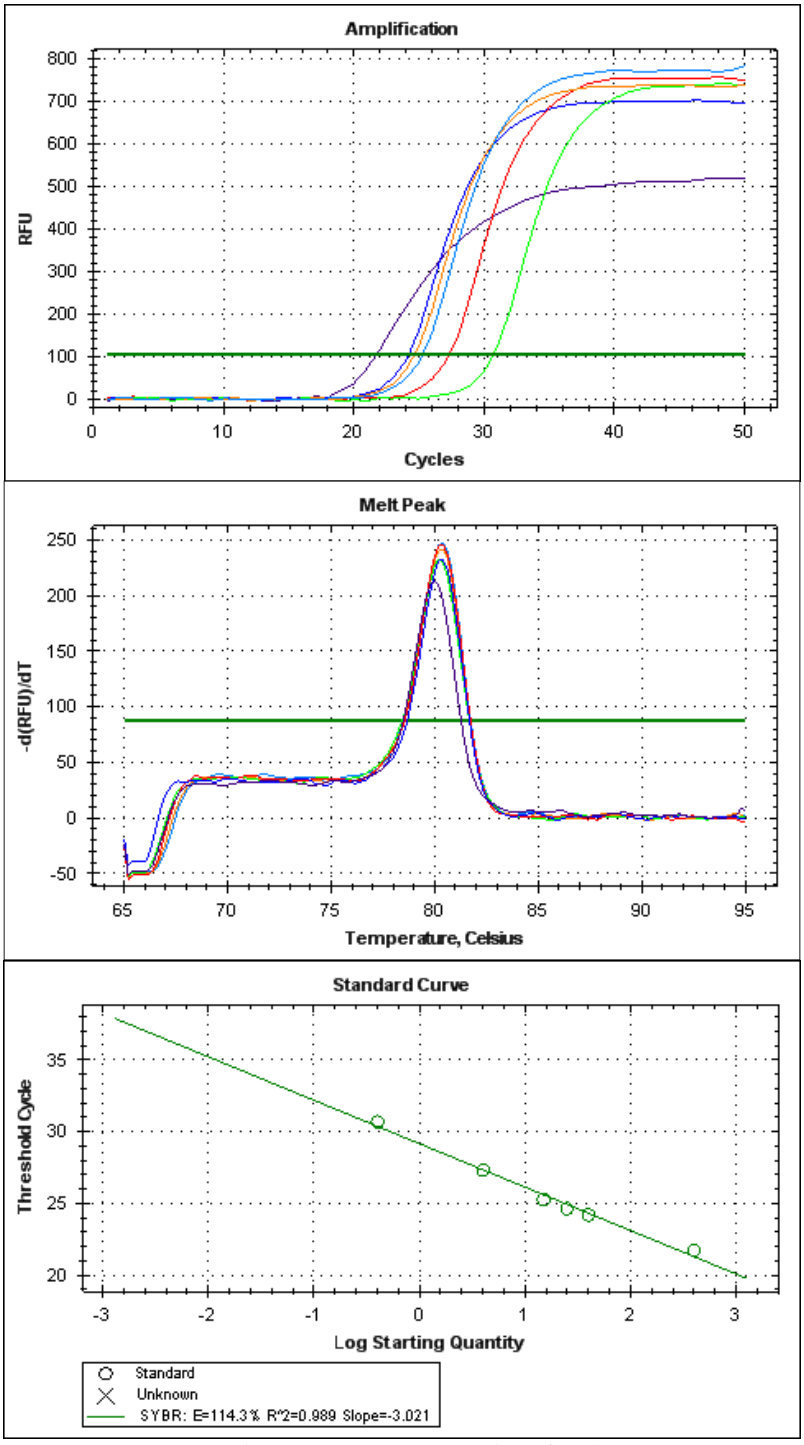
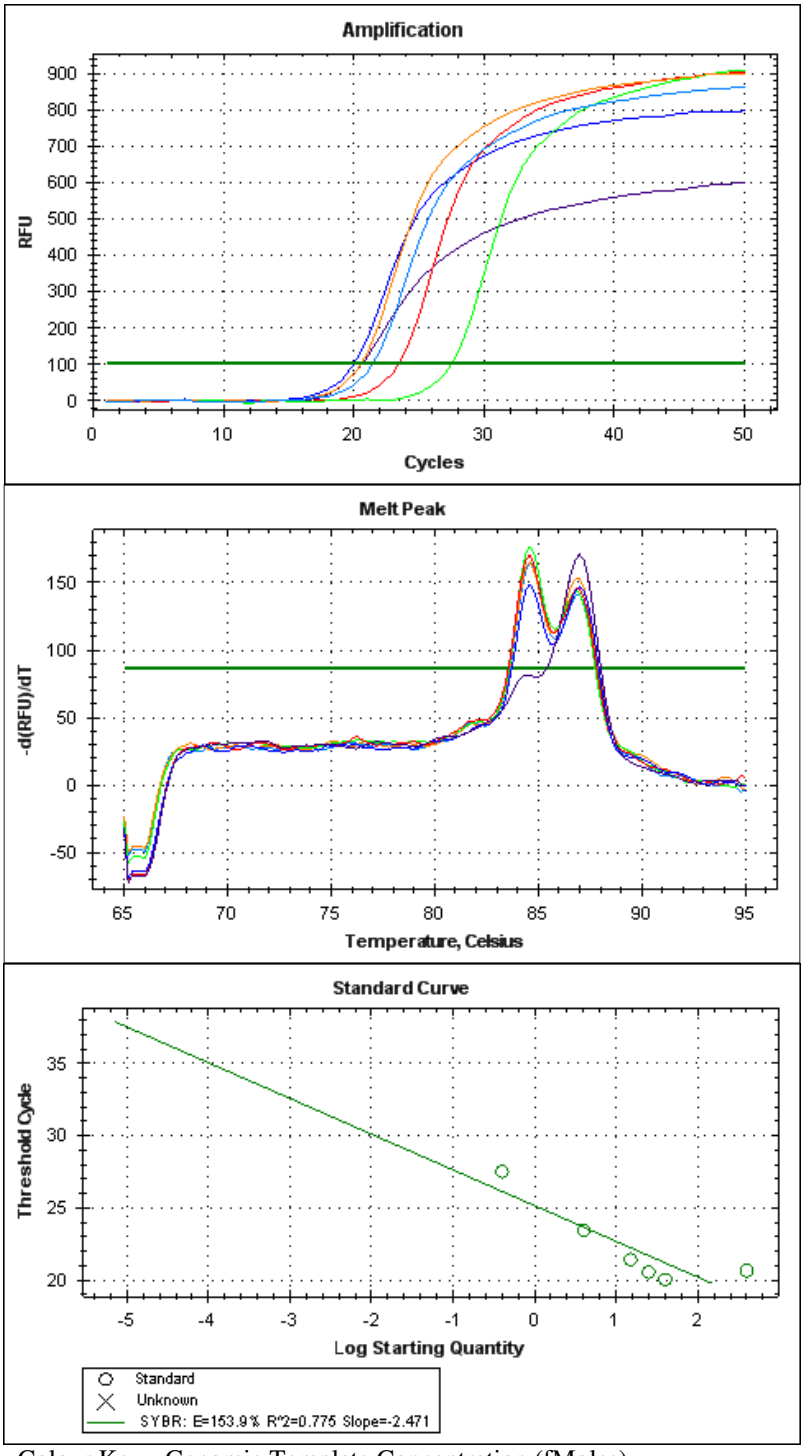


Figure 47: IGF1_8 Amplification Curve, Melt Peak and Standard Curve. IGF1_8 demonstrates poor efficiency over this dynamic range ($6.16x10^{1}$-$6.16x10^{-2}$fM) with E = 114.3%, slope = -3.021 and r^2 = 0.989. Sample melt peaks indicate single target amplification. Amplification curves display a sigmoid shape but do not follow the anticipated pattern of 3.3 cycles between 10-fold serial dilutions for higher $6.16x10^{1}$fM target samples. Agarose gel electrophoresis indicates that target specific amplification was achieved for all 6 samples.



Colour Key - Genomic Template Concentration (fMoles)

$6.16x10^{1}$	$6.16x10^{-1}$	3.85
6.16	$6.16x10^{-2}$	2.31

Figure 48: IGF1_9 Amplification Curve, Melt Peak and Standard Curve. IGF1_9 demonstrate poor efficiency and specificity over this dynamic range (6.16×10^{1}-6.16×10^{-2}fM) with E = 152.9%, slope = -2.471 and r^2 = 0.775. Amplification curves display a sigmoid shape however adjacent curves do not follow the anticipated pattern of 3.3 cycles between 10-fold serial dilutions. All sample melt peaks indicate equimolar duplex amplification bar the largest $6.16x10^{1}$fM target sample which displays positively skewed amplification of the target amplicon. This poor specificity was confirmed by agarose gel electrophoresis.

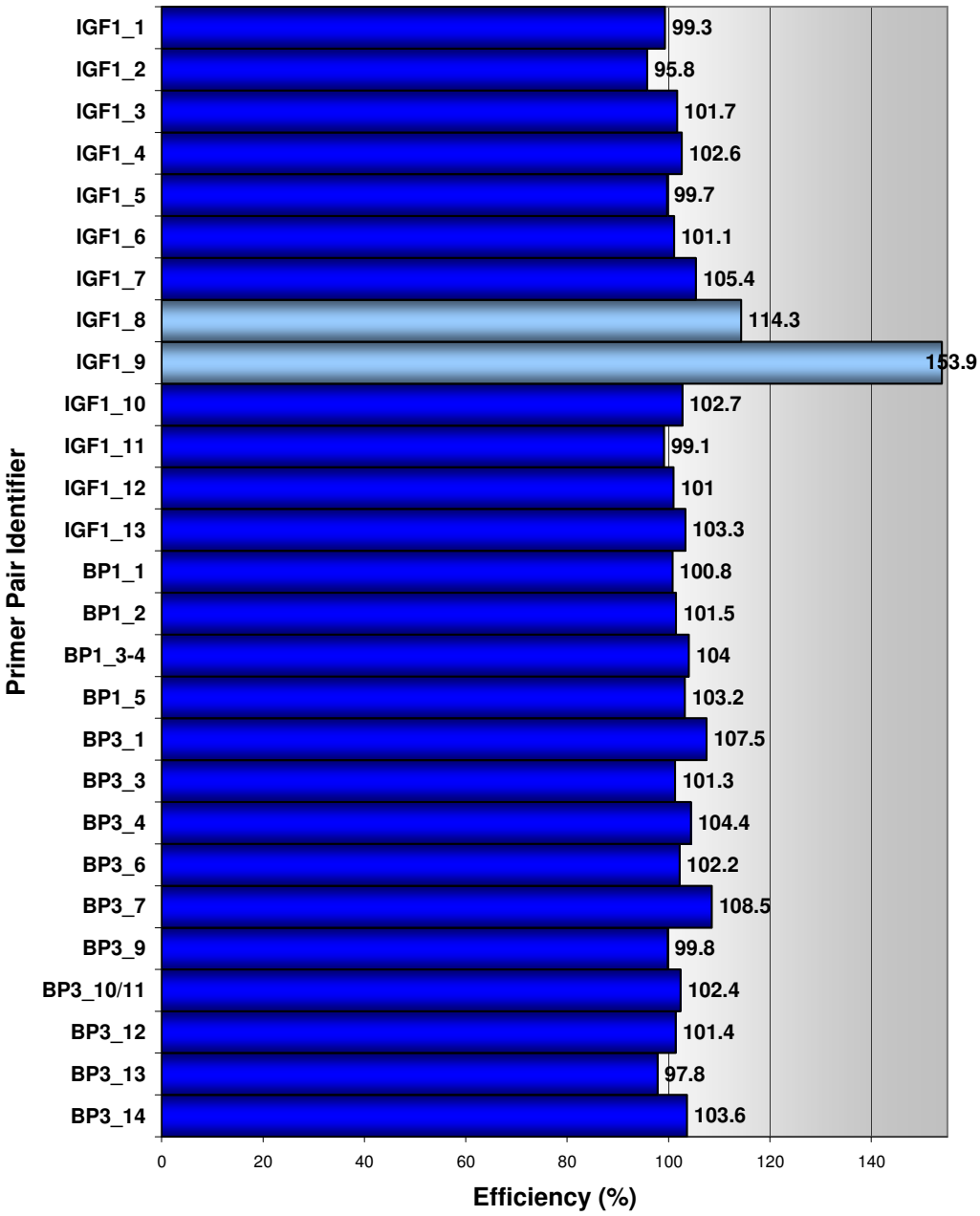


Figure 49: RT-PCR Derived Efficiencies for Singleplex Primer Pairs. RT-PCR was carried out using uniform conditions as described (4.2.7 Efficiency / Dynamic Range Determination) for all primer pair and efficiencies plotted against identifiers. Primer pairs IGF1_8 and IGF1_9, with efficiencies of 114.3 and 153.9% respectively, exceed the recommended 90-110% efficiency boundaries indicating poor primer pair performance. All other primer pairs are within the recommended efficiency range with an arithmetic mean of 102.00, standard deviation of 2.79 and variance of 7.78. Despite receiving good efficiency, slope and r^2 scores IGF1_10 and BP3_14 (along with poorly efficiency IGF1_9) also displayed poor specificity upon agarose gel electrophoretic analysis.

Table 32 : Table displaying efficiencies, r^2, slope and quality designation derived by RT-PCR analysis of singleplex primer pairs. Efficiency is a measure of primer pair performance and should range between 90-110% for primer pairs with good functionality. The slope of the standard curve is directly related to the average efficiency of amplification and should be between -3.6 and -3.1 while r2 (correlation coefficient) is indicative of the quality of the fit of the data points plotted to the standard curve, r2 should be >0.95. Primer pairs with scores exceeding set limits in any categories (i.e. IGF1_8 and IGF1_9) are determined to display "poor" efficiency.

	Primer Identifier	Efficiency	r^2	Slope	Functionality
1	IGF1_1	99.3	0.986	-3.338	Good
2	IGF1_2	95.8	0.988	-3.426	Good
3	IGF1_3	101.7	0.988	-3.282	Good
4	IGF1_4	102.6	0.991	-3.261	Good
5	IGF1_5	99.7	0.993	-3.329	Good
6	IGF1_6	101.1	0.991	-3.295	Good
7	IGF1_7	105.4	0.991	-3.199	Good
8	IGF1_8	114.3	0.989	-3.021	Poor
9	IGF1_9	153.9	0.775	-2.478	Poor
10	IGF1_10	102.7	0.980	-3.260	Good
11	IGF1_11	99.1	0.990	-3.344	Good
12	IGF1_12	101.0	0.963	-3.297	Good
13	IGF1_13	103.3	0.992	-3.244	Good
14	BP1_1	100.8	0.984	-3.302	Good
15	BP1_2	101.5	0.992	-3.285	Good
16	BP1_3/4	104.0	0.985	-3.230	Good
17	BP1_5	103.2	0.978	-3.247	Good
18	BP3_1	107.5	0.993	-3.155	Good
19	BP3_3	101.3	0.985	-3.291	Good
20	BP3_4	104.4	0.983	-3.222	Good
21	BP3_6	102.2	0.994	-3.270	Good
22	BP3_7	108.5	0.992	-3.134	Good
23	BP3_9	99.8	0.994	-3.326	Good
24	BP3_10/11	102.4	0.986	-3.266	Good
25	BP3_12	101.4	0.988	-3.289	Good
26	BP3_13	97.8	0.998	-3.376	Good
27	BP3_14	103.6	0.992	-3.239	Good

4.3.5 Multiplex PCR Size Determination

The Experion microfluidic electrophoresis platform was used for separation and quantitation of multiplex PCR products. The resolution capacity was tested and high functionality primer pairs (as identified by preliminary singleplex PCR and RT-PCR analysis) separated into groups. Reaction components and thermocycle profile adjustments were optimised for each multiplex reaction to facilitate high-specificity, suitably equimolar amplification of multiple targets. Four multiplex reactions amplifying 14, six, five and two targets, capturing variation from 29 bins across candidate genes IGF1_1, IGFBP1 and IGFBP3 were constructed. Experion-generated electropherograms are displayed for 14-ples, 6-plex, 5-plex and 2-ples reactions in Figure 50, Figure 51, Figure 52 and Figure 53 respectively. Each electropherogram peak is labelled with its corresponding primer-pair identifier and Experion-estimated amplicon size. A virtual gel corresponding to the sample is displayed below each graph. Amplicon size data displayed in table format may be accessed in electronic appendix Table 13, Table 14, Table 15 and Table 16.

Amplification of all 14, six, five and two targets was successful. The 14-plex amplified the largest amplicon size range with targets spanning 428bp from 72-500bps. 6-plex amplicons ranged from 154-248bps, 5-plex ranged from 156-250bps and the 2-plex amplified targets of 147 and 260bp. The size designations made exceeded the anticipated size in most cases (+ ~12bp).

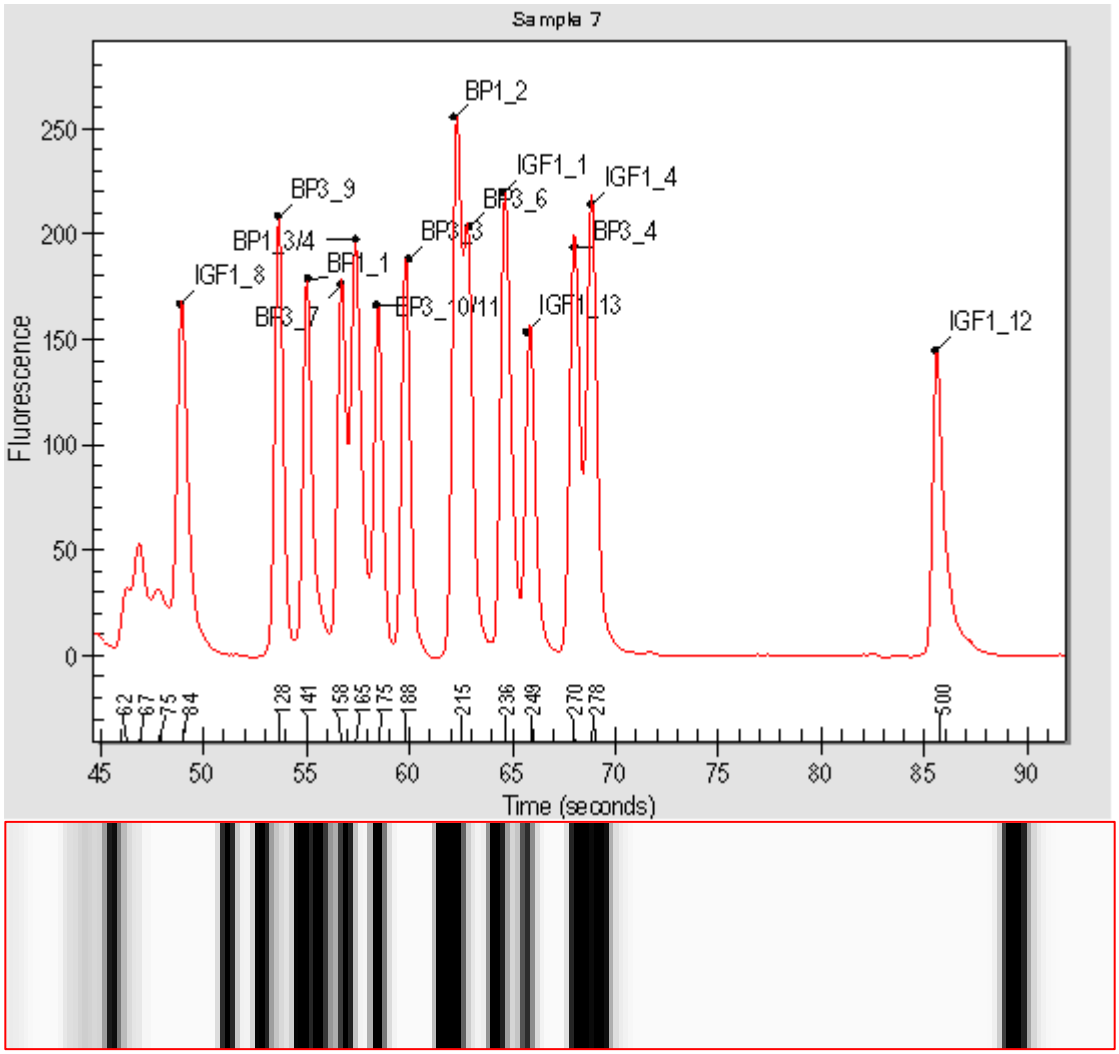


Figure 50: Multiplex (14-plex) PCR amplification of 14 primer pair targets ranging in size from 72-500bps was performed using the Qiagen multiplex PCR kit. Using optimised reaction mix and thermocycle profiles 14 targets of anticipated size range were produced (+/- 12bp).

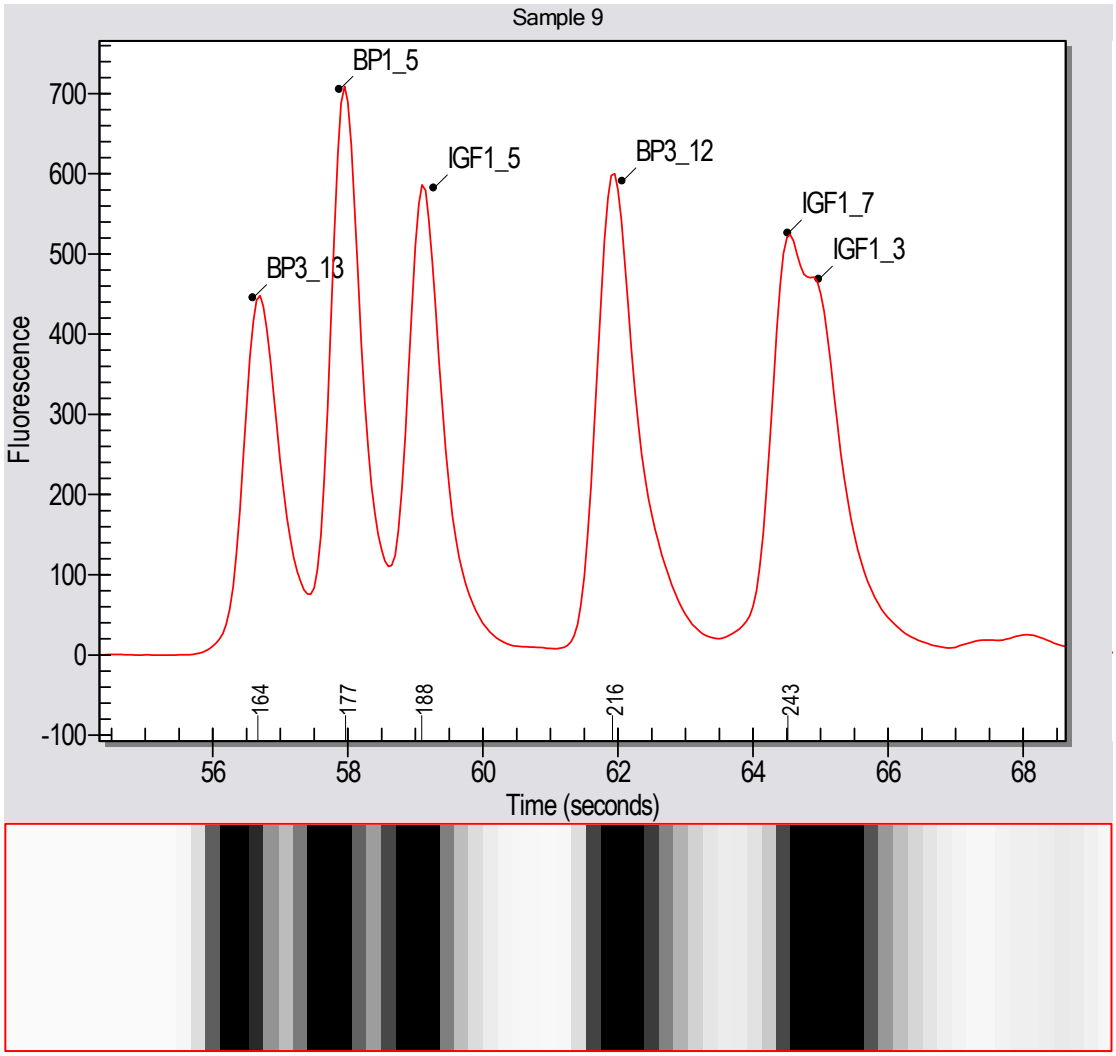


Figure 51: Multiplex (6-plex) PCR amplification of 6 primer pair targets ranging in size from 154-248bps was performed using the Qiagen multiplex PCR kit. Using optimised reaction mix and thermocycle profiles 6 targets of anticipated size range were produced (+/- 13bp).

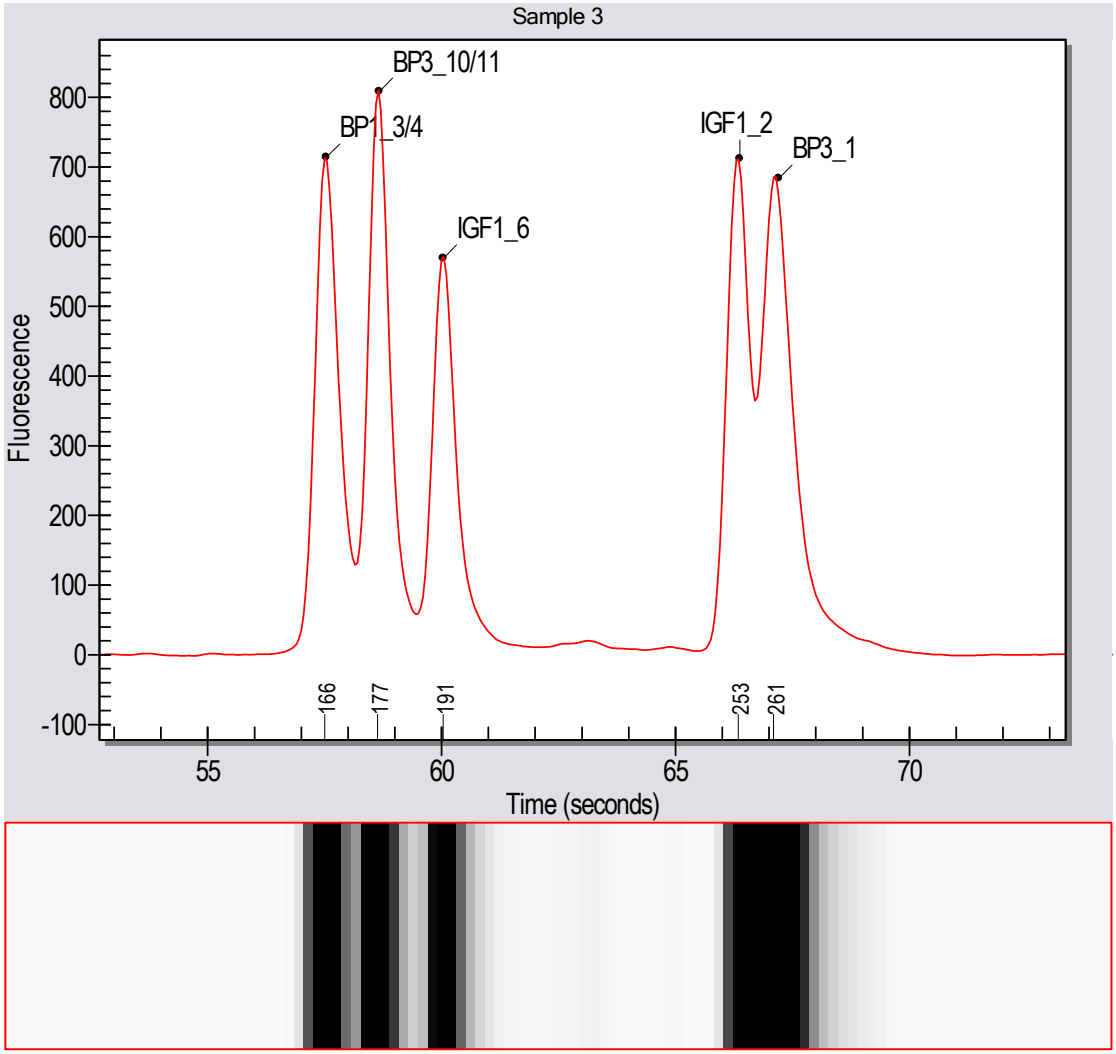


Figure 52: Multiplex (5-plex) PCR amplification of 5 primer pair targets ranging in size from 156-250bps was performed using the Qiagen multiplex PCR kit. Using optimised reaction mix and thermocycle profiles 6 targets of anticipated size range were produced (+/- 17bp).

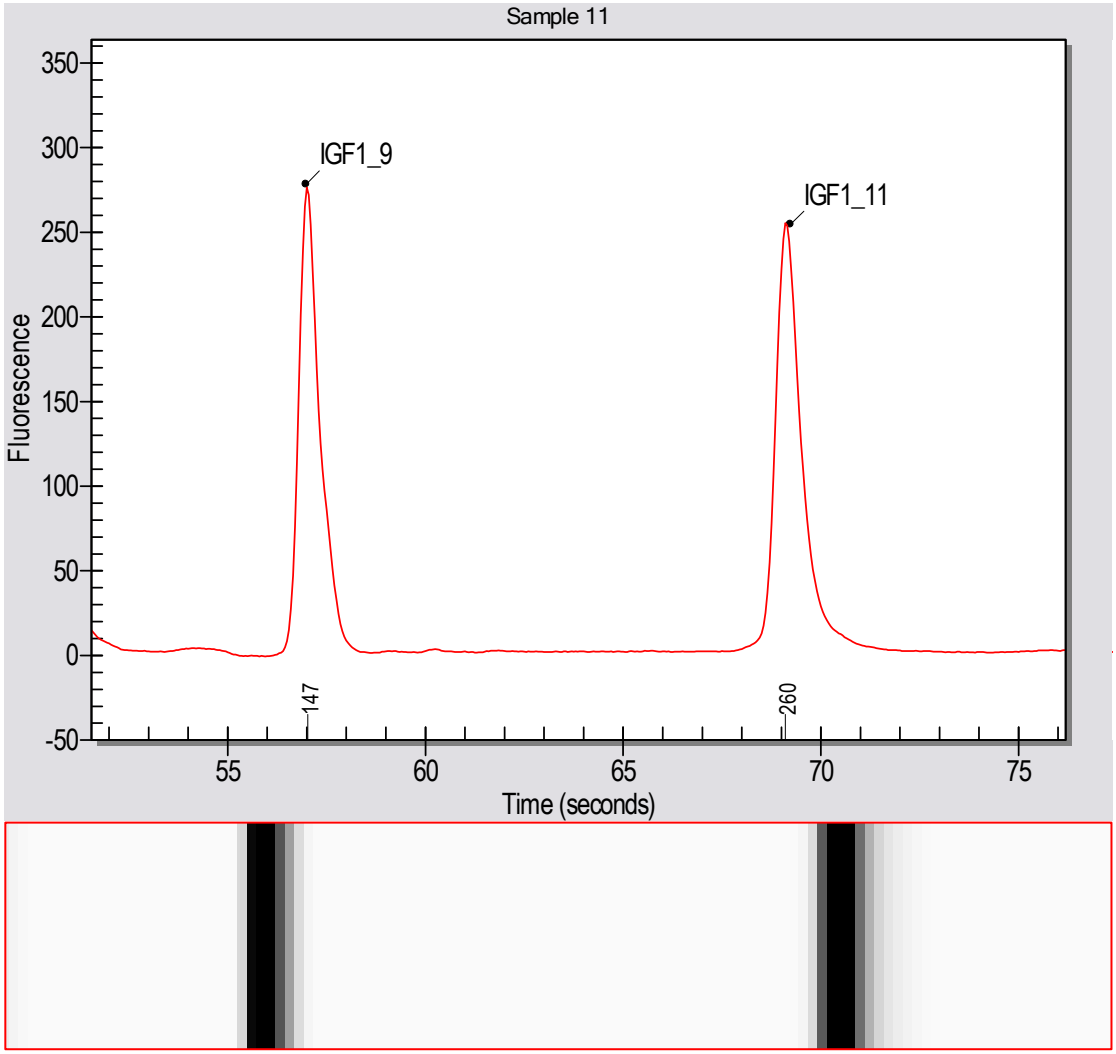


Figure 53: Multiplex (2-plex) PCR amplification of IGF1_9 (141bp) and IGF1_11 (248bp) was performed using Ampli*Taq* Gold and thermocycle profile as described (Table 29: 2-plex Thermal Cycle Profile. Use of this high processivity time-release polymerase was implemented to reduce non-specific amplification of repeat region primer pair IGF1_9 which performed poorly using alternate polymerases/protocols. The final reaction and thermocycle profile was found to facilitate amplification of both targets (+/- 12bp).

4.3.6 Multiplex PCR Yield Determination

Multiplex primer pair yields pertaining to 14-plex, 6-plex, 5-plex and 2-plex figures displayed previously (Figure 50, Figure 51, Figure 52 and Figure 53) determined by comparative analysis against molecular weight markers (Experion facilitated quantitation) are displayed in column charts. Each column is labelled with its primer pair identifier and corresponding yield (ng/µl). Data displayed in table format may be accessed in electronic appendix Table 17, Table 18, Table 19 and Table 20 for 14-plex, 6-plex, 5-plex and 2-plex reactions respectively.

All reactions produced strong product yields which displayed relatively equimolar profiles. The 14-plex reaction displayed a mean yield of 6.06ng/µl ranging from 4.36 - 8.24ng/µl for primer pairs IGF1_13 and IGF1_8. The 6-plex and 5-plex reactions produced increased mean yields of 10.07ng/µl and 10.57ng/µl ranging from 7.26 - 12.45 (IGF1_7 and BP3_12) and 8.82-12.31ng/µl (IGF1_2 and BP3_) respectively. The 2-plex produced the highest mean yield of 11.44ng/µl and most equimolar amplification of targets IGF1_11 and IGF1_9 (10.48 and 12.4ng/µl respectively).

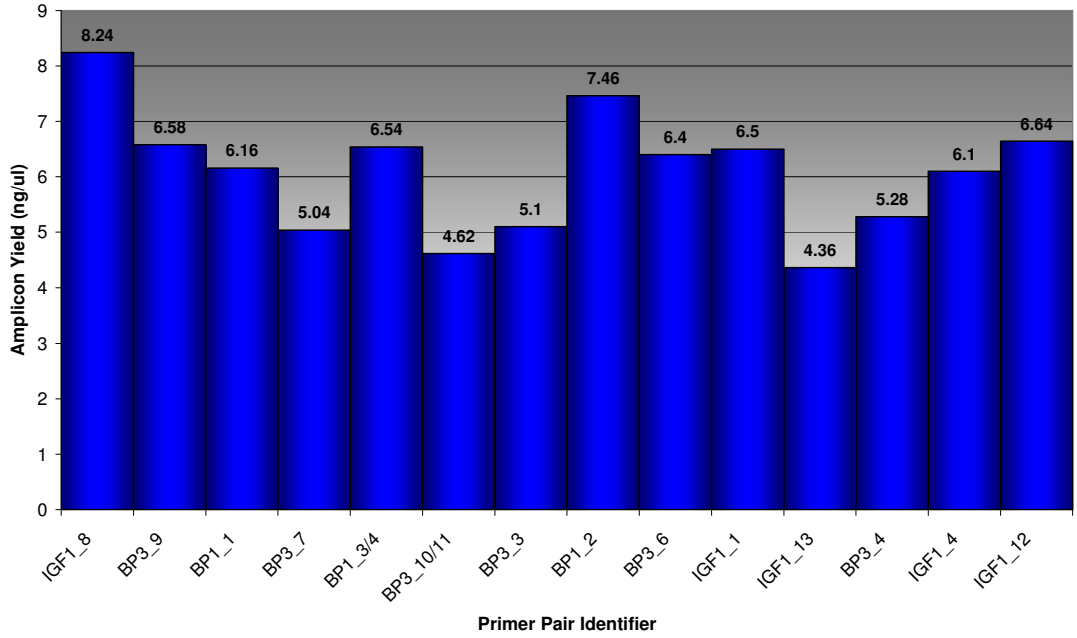


Figure 54: Multiplex (14-plex) PCR target yields for 14 primer pairs derived using optimised 14-plex protocol and quantified using the Experion microfluidic electrophoresis and detection system. Using optimised reaction mix, adjusted relative primer pair concentrations and thermocycle profiles as described suitably-equimolar amplification of all targets was achieved; resulting in a mean amplification yield of 6.06ng/µl for all targets, ranging from 4.36 - 8.24ng/µl for primer pairs IGF1_13 and IGF1_8 respectively.

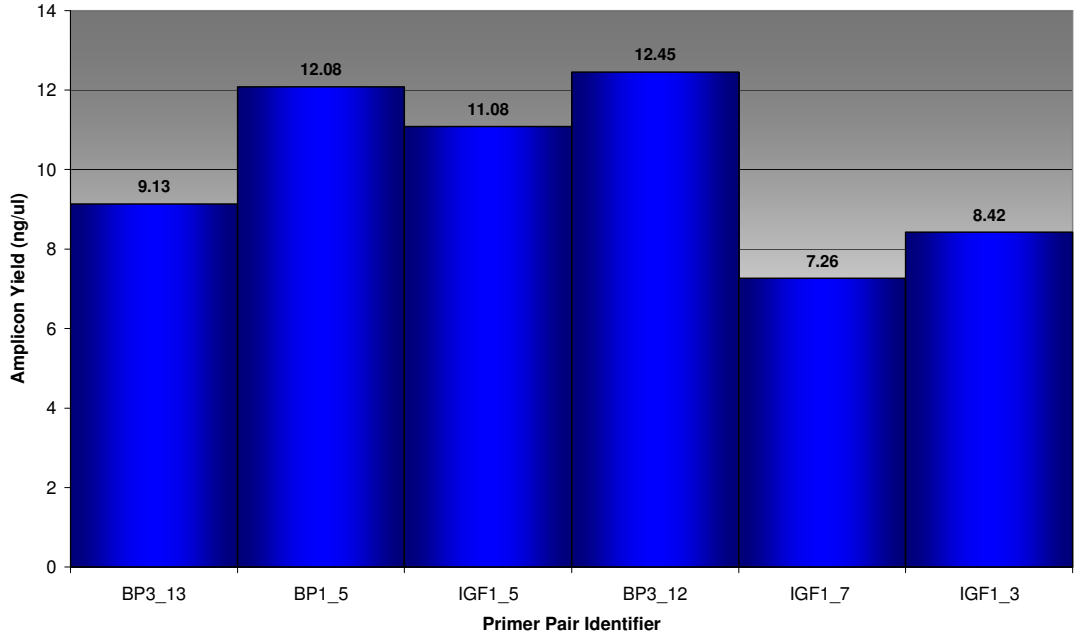


Figure 55: Multiplex (6-plex) PCR target yields for 6 primer pairs derived using optimised 6-plex protocol and quantified using the Experion microfluidic electrophoresis and detection system. Optimisation of both reaction mix and thermocycle profiles were used to facilitate reasonably equimolar amplification of all targets. Relative primer pair adjustment was used to adjust poorer efficiency resulting in relatively equimolar amplification of all primer pairs, with a mean yield of 10.07ng/µl, ranging from 7.26 - 12.45 for primer pairs IGF1_7 and BP3_12 respectively.

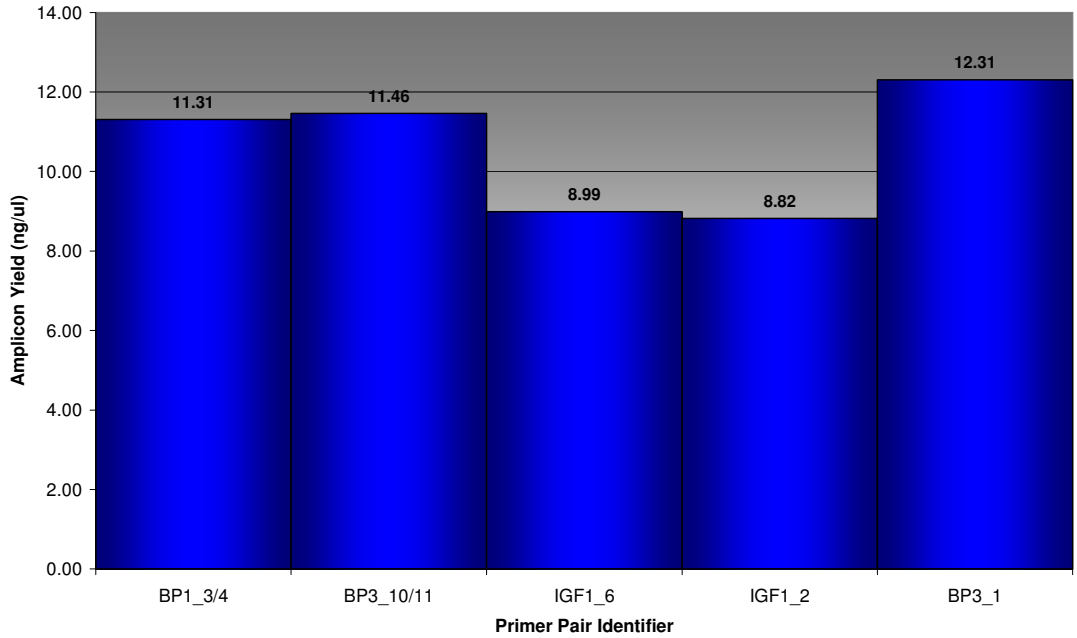


Figure 56: Multiplex (5-plex) PCR target yields for 5 primer pairs derived using optimised 5-plex protocol and quantified using the Experion microfluidic electrophoresis and detection system. Optimal thermal-cycle protocol and relative primer pair adjustment was used to facilitate suitably-equimolar amplification of all primer pairs, resulting in a mean amplification yield of 10.57ng/µl for all targets, ranging from 8.82-12.31ng/µl for primer pairs IGF1_2 and BP3_1 respectively.

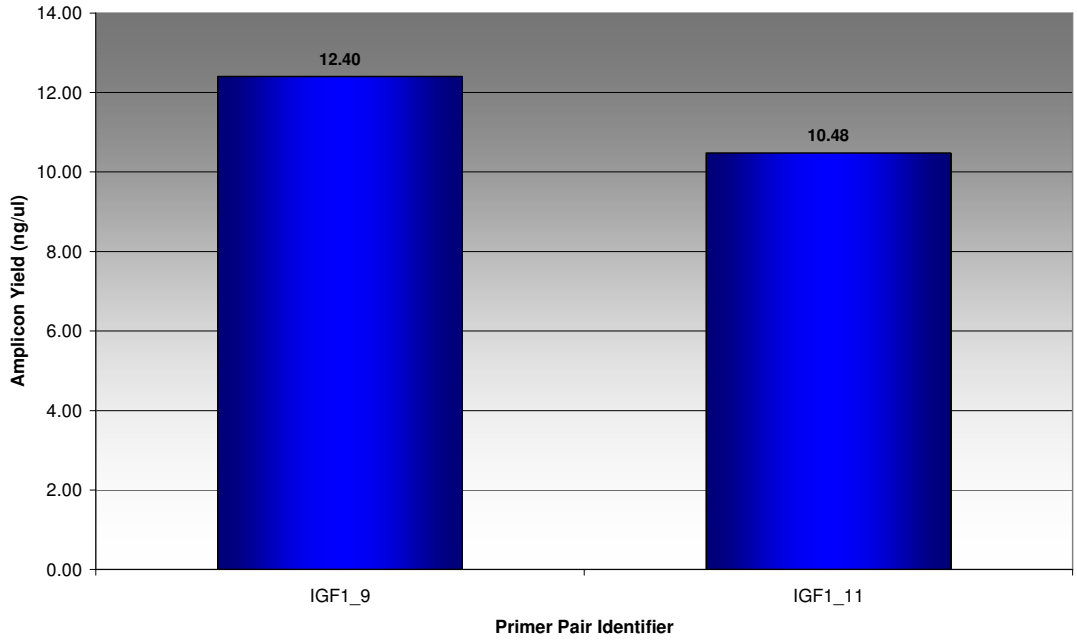


Figure 57: Multiplex (2-plex) PCR target yields for two primer pairs derived using optimised 2-plex protocol and quantified using the Experion microfluidic electrophoresis and detection system. A mean yield of 11.44ng/µl was determined, ranging from 10.48 – 12.4ng/µl for primer pairs IGF1_11 and IGF1_9 respectively.

4.3.7 Relative Multiplex End-point Efficiencies

The concentration of primer pairs used for multiplex amplification was not equal; as such relative amplification efficiency (accounting for initial primer pair concentration and resultant yield) was determined as described (see 4.2.10 Experion Electrophoresis) and relative end-point efficiencies plotted for each multiplex reaction.

The 14-plex reaction contained the widest range of amplicon sizes ranging from 72-500bp for IGF1_8 and IGF1_12 respectively (Figure 58). Efficiency appears to have been influenced by amplicon size within this 14-plex format with smallest (<160bp) and largest (500bp) amplicons suffering most significantly. Within the 6-plex reaction both primer pairs IGF1_5 and BP3_13 displayed highest end-point amplification efficiency approximately 3 times that of other amplicons in the set, a significantly increased relative efficiency was also seen for primer pair BP3_10/11 within the 5-plex reaction (Figure 59 and respectively Figure 60). The reasons for increased efficiency in all three cases is unknown; neither innate primer features or characteristics determined during singleplex optimisation assays indicated a tendency toward superior performance of these pairs relative to others in their respective sets. The smallest 2-plex reaction displayed relatively equal efficiencies for both primer pairs (Figure 61).

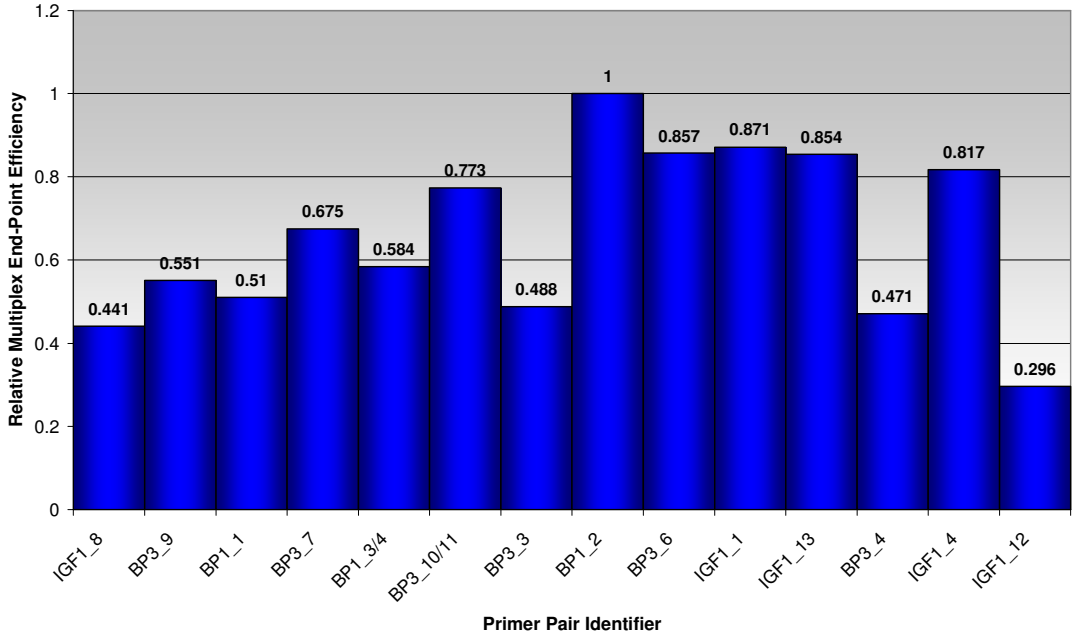


Figure 58: Relative, end-point primer pair efficiencies within the optimised 14-plex format previously described. Average relative efficiency of 60.73% was determined with a standard deviation of 23.09% and variance of 5.33%. Efficiency appears to be influenced by amplicon size within this 14-plex format where amplicons range from 72-500bp, with smaller (<160bp) and largest (500bp) amplicons suffering most significantly.

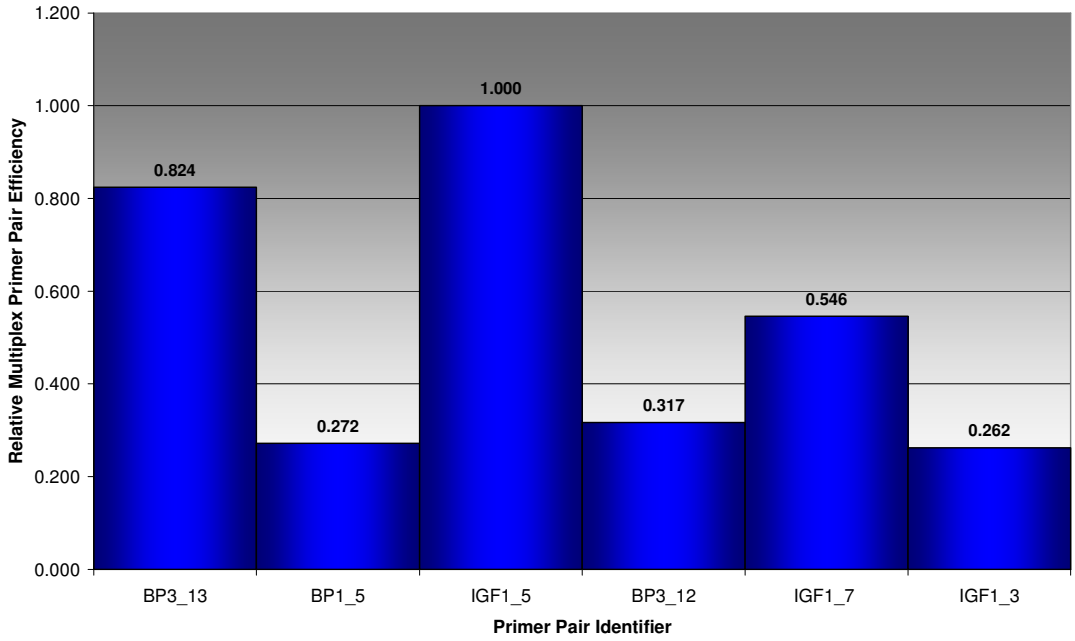


Figure 59: Relative, end-point primer pair efficiencies within the optimised 6-plex format previously described. Average relative efficiency of 53.68%was determined with standard deviation of 31.33 and variance of 9.82%. Both primer pairs IGF1_5 and BP3_13 displayed highest end-point amplification efficiency in this format.

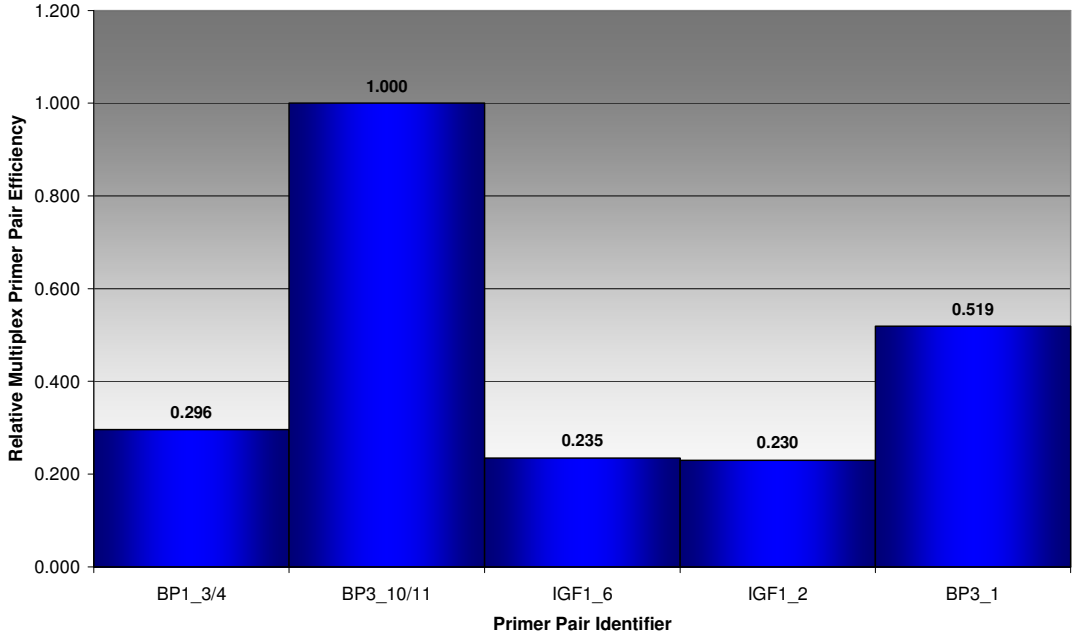


Figure 60: Relative, end-point primer pair efficiencies within the optimised 5-plex format previously described. Average relative efficiency of 45.60% was determined with standard deviation of 32.61% and variance of 10.63%. Primer pair BP3_10/11 displayed significantly higher efficiency than all other pairs in this set.

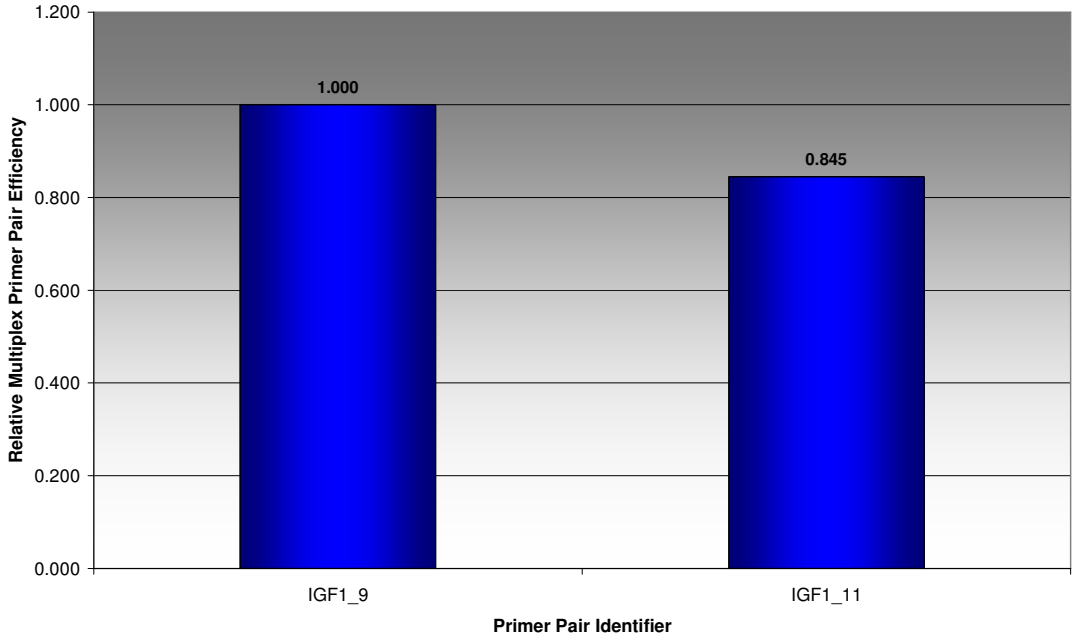


Figure 61: Relative, end-point primer pair efficiencies within the optimised 2-plex format previously described. Average relative efficiency of 92.25% was determined with standard deviation of 10.96% and variance of 1.20%. Relative 2-plex efficiency mirrored that displayed by Ampli*Taq* gold singleplex amplification of both pairs.

4.3.8 14 Multiplex Optimisation

Amplification of the largest 14-plex set required extensive optimisation. Annealing time adjustment was required to adjust for the wider size range of target amplicons and factors affecting reactions kinetics including total primer concentration and total reaction volumes were investigated.

Longer annealing times of 3mins were found to favour 14-plex amplification efficiency of larger amplicons (most notable for the largest 500bp IGF1_12) (Figure 58). While a use of alternate total primer concentrations ranging from 3.0 and 1.0µM were found to effect amplification efficiency in a size dependant manner (Figure 59). A similar effect was seen with use of alternate total reaction volumes (25-100µl). A balance between multiple contributory factors was therefore required to produce suitably equimolar amplification of all targets.

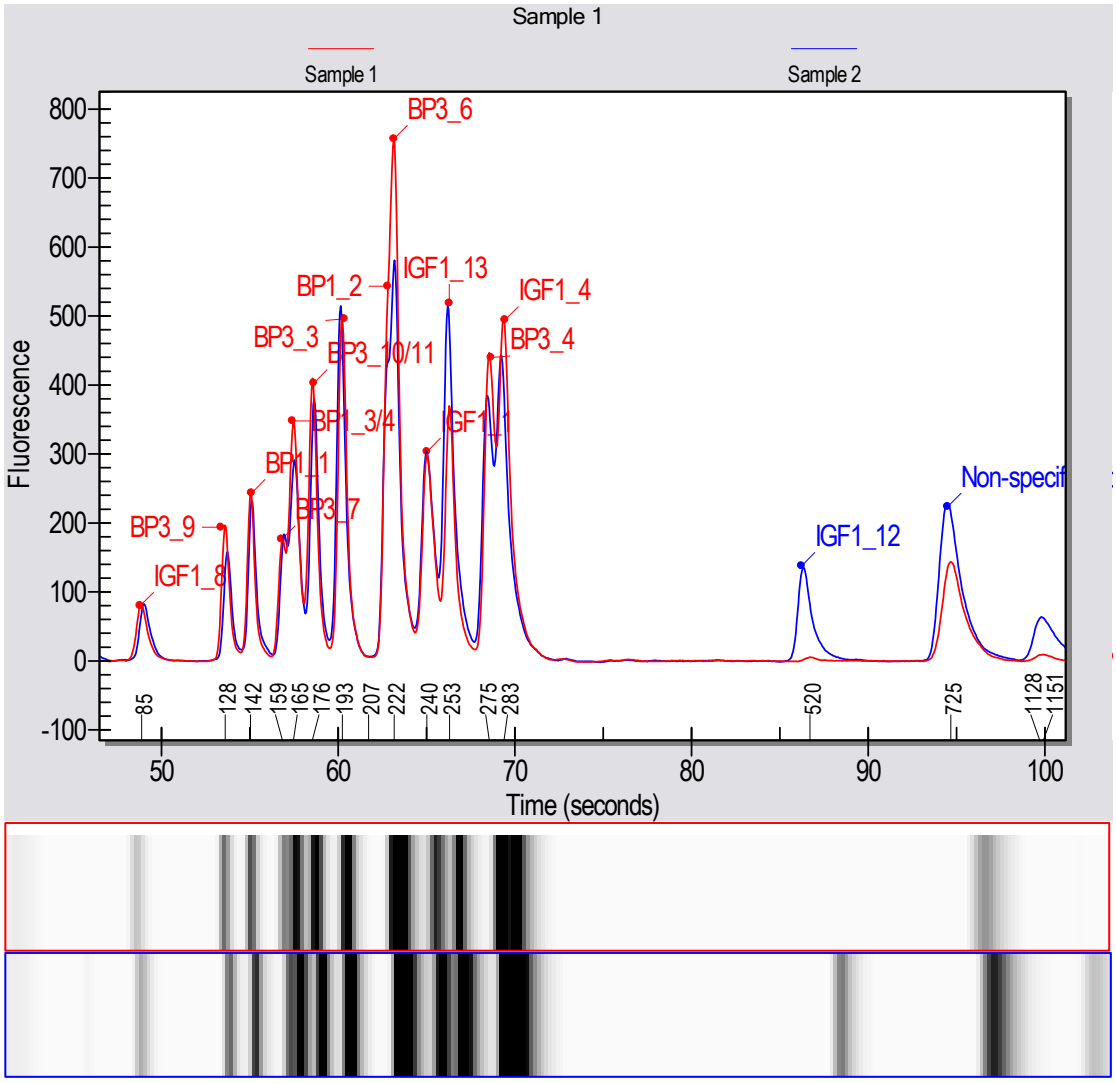


Figure 62: Effect of Annealing time on Amplification of larger Amplicons. 14-plex samples 1 and 2 were amplified using annealing times of 1min30s and 3mins respectively. An increase in amplification efficiency can generally be seen for larger amplicons (>250bp), this effect is most apparent with largest 500bp IGF1_12 amplicon, which is present at a concentration of 2ng/µl in 3min annealing sample 2, but absent from sample one.

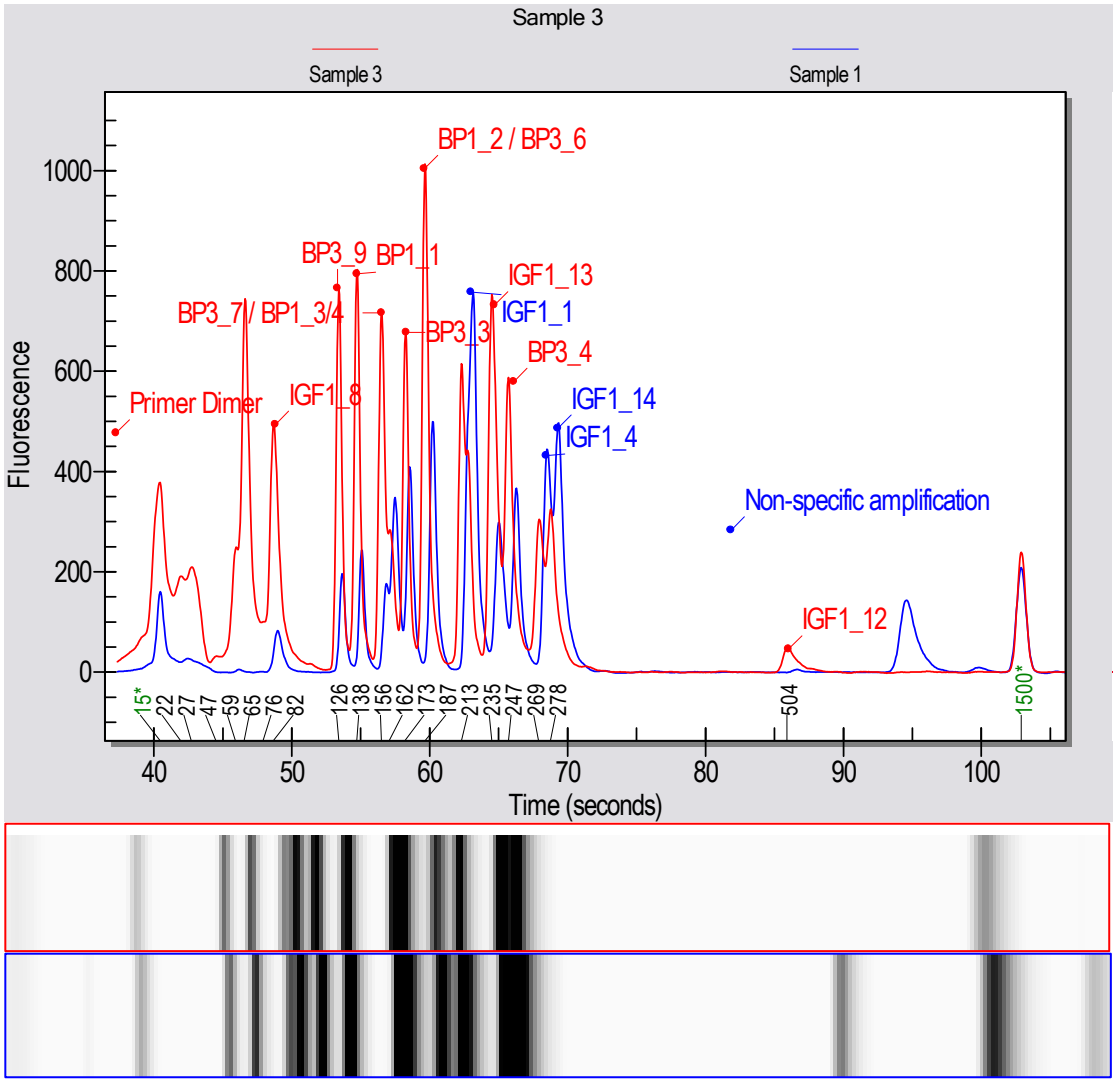


Figure 63: Effect of total primer concentration on multiplex amplification. 14-plex amplification was performed using an equimolar primer mix at a final concentration of 3.0 and 1.0µM for sample 3 and 1 (highlighted in red and navy) respectively. Use of higher total primer concentration results in preferential amplification of smaller amplicons, small non-specific amplicons and primer dimers while use of lower total primer concentration results in preferential amplification of larger amplicons and larger non-specific products.

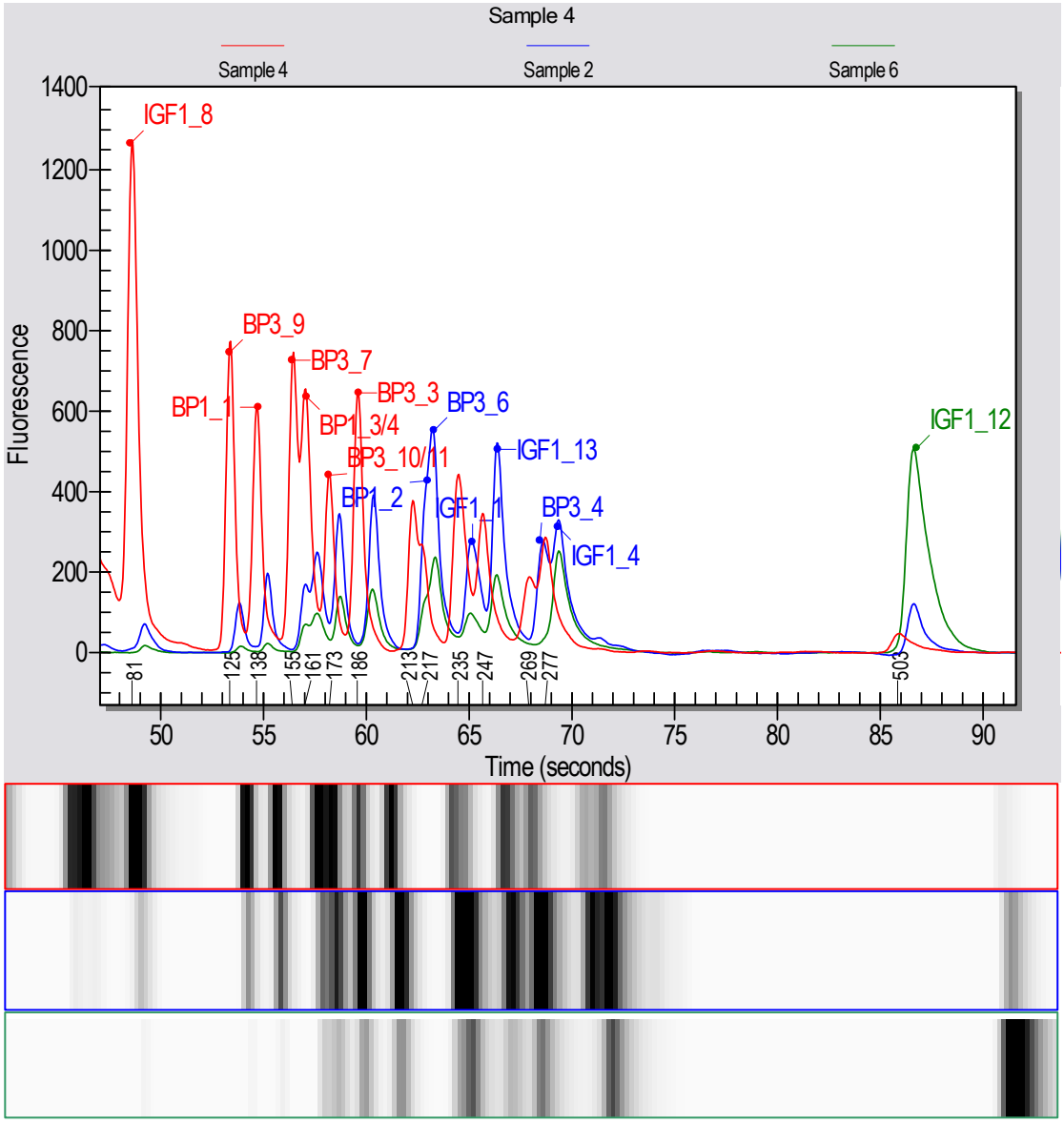


Figure 64: Effect of total reaction volume on 14-plex amplification using alternate reaction volumes from 25-100µl with common reaction conditions and reactant concentrations as specified. Increased volumes result in reduced speed of heat transfer within the reaction solution, as such "true" thermocycle segment times may be impacted and amplification effected. Using the reaction conditions specified 50µl total reaction volume (blue trace) resulted in most equimolar amplification of the target size range used in this reaction mix. Smaller 25µl (green trace) and larger 100µl reaction volumes (red trace) resulted in preferential amplification of larger and smaller targets respectively.

4.4 Discussion

Multiplex PCR is a complex reaction that often results in low product yield, amplification failure and or non-specific amplification. Complications of this nature may be influenced by incompatibilities with regard to annealing temperature, primer cross-reactivity and divergent amplification efficiencies. Increasing reaction complexity by use of higher-dimensionality multiplex formats often results in exasperation of these effects. It is therefore essential to perform extensive analysis and optimisation of primer pairs using both singleplex and multiplex formats if robust amplification of this nature is to be performed [180,213,219]. The following section discusses the results of experimental PCR evaluation including singleplex end-point and real-time analysis as well as discussion of multiplex optimisation and the final PCR profiles.

4.4.1 Single-plex End-point PCR

All primer pairs were initially tested in singleplex using end-point PCR to assess their specificity. Using the Qiagen protocol as previously described (see 4.2.3 Single-plex End-Point PCR); 24 of the 27 primer pairs displayed target specific amplification, however three primer pairs IGF1_9, IGF1_10 and BP3_14 located in or amplifying repeat region targets displayed non-specific amplification. The Qiagen kit comes with a pre-blended master mix optimised for multiplex amplification of standard targets and contains a number of moieties which may act to reduce specificity when used for amplification of repeat regions. These include a high final $MgCl_2$ concentration of 3mM and MP-factor. Metal ions (particularly divalent cations such as Mg^{2+}) act to stabilise hydrogen bond formation between single stranded nucleic acid sequences [220]. This can be useful for increasing yields however it can also act to stabilise bonding of primers with

non-complementary repeat regions or formation of amplicon-constituent hairpins leading to non-specific amplification as was seen in this instance [170,180,183]. Further optimisation was required therefore, to ameliorate repeat region pairs IGF1_9, IGF1_10 and BP3_14. An Ampli*Taq* no-$MgCl_2$ protocol as described (see 4.2.3 Single-plex End-Point PCR) was used to achieve these ends. This protocol uses a higher processivity Ampli*Taq* DNA polymerase which should facilitate faster elongation and higher fidelity [190]. The removal of $MgCl_2$ should also act to reduce secondary structure formation and potential deletion mutagenesis artifact formation. Amplification specificity was again good for all 24 non-repeat region primer pairs. Specificity was found to improve somewhat for all problematic pairs however IGF1_9 benefitted most significantly from this treatment with one single target amplicon (~155bp) produced in this instance.

A wide range of adjuvants including DMSO and Q-solution have been developed to reduce hairpin formation while allowing conditions conducive to hybridisation to be retained (formamide and Dimethyl sulfoxide) [182]. Time-release methods which allow slow-activation of polymerase have also been shown to reduce excess-polymerase induced non-specific amplification. While high-fidelity polymerases are also known to reduce aberrant amplification. Six alternate treatment types (A-F inclusive) employing a number of these approaches were used to try to reduce non-specific amplification of repeat region targets IGF1_10 and BP3_14. $MgCl_2$ restricted manual hot start with standard *Taq* (treatment A) and time-release Ampli*Taq* Gold facilitated amplification with and without DMSO destabilising agent (treatment B, C and F; 0%, 5% and 10% DMSO respectively) were used, as was the Qiagen 3mM $MgCl_2$ constituent mix with and

without Q-solution isostabilizing agent (treatment D and E) (see 4.2.4 Repeat Region Amplification Optimisation for full details).

IGF1_10 amplification produced non-specific artefacts for all treatment types. As expected use of manual hot start, standard *Taq* (treatment A) resulted in poorest specificity with nine amplification bands ranging in size from ~187-1,027bp produced. The Qiagen master 3mM $MgCl_2$ protocol (treatment D) resulted in production of five amplification bands ranging in size from 658 – 199bp. Specificity was increased by addition of 5% isostabilizing Q-solution (treatment E), with three bands of 246, 207 and 199bp produced in this instance. Use of the no- $MgCl_2$, time-release Ampli*Taq* gold protocol (treatments B, C and F) again resulted in amplification of three bands in all instances. The relative proportions of putative 199bp target did significantly increase with increasing DMSO concentration however and addition of 10% DMSO (treatment F) produced a putative target concentration, approximately twice that of non-specific artefacts.

As previously discussed; PCR tends to be less specific and efficient when primers containing 3'adenine or thymine terminal and/or penultimate nucleotides are used [136]. Target constraints meant that IGF1_10 primers were designed without a G/C clamp on either primer with forward or reverse primers. Loss of this GC clamping feature, particularly in a repeat region already prone to mispriming[10], is likely to have impacted upon specificity as was seen in this instance.

[10] Mispriming occurs due to binding of primers to unintended template sites, subsequent amplification results in the formation of non-specific product

Target specific amplification of BP3_14 was also quite challenging; three amplicons (289, 269 and 258bp) were produced for treatments A, B, C, D and E. Addition of 10% DMSO (treatment F) increased relative concentration of doublet 289 and 269bp bands however yields were heavily impacted and a total yield of just 4.44ng/µl was achieved. As previously discussed high-specificity doublet formation can result from indel heterogeneity. Five indels are known to exist in this BP3_14 target region however at 289bp the secondary band is outside the predicted indel variable range (274- 293bp) and as such doublet formation is more likely to be attributable to deletion mutagenesis or non-specific amplification.

4.4.2 Single-plex Real-Time PCR

4.4.2.1 Primer Pair Annealing Temperature (T_A)

Annealing Temperature uniformity is one of the most critical requirements for successful multiplex PCR amplification [132]. According to Panjkovich T_A prediction; all primers in the final IGF multiplex set should have an optimal T_As of between 64.47 and 65.81°C (mean T_A = 65.24°C), however prediction agreement is strongest for oligonucleotide sequences <20–22 nucleotides in length which have a CG content of 40-60%. All primers designed during this study fall within recommended GC limits (40-59%), however with a mean primer length of 27 nucleotides (ranging from 22-30bases), primers exceed high-specificity prediction limits. As such experimental assessment of optimal T_A was especially required. Five temperatures ranging from 59.9-68.8°C, spanning the anticipated optimal annealing (~65°C) were tested for each primer pair. Amplification curve, melt peaks and relative yields were plotted and assessed to determine optimal experimental T_A.

Manual evaluation of amplification curves was used to assess whether the reaction followed the expected sigmoid amplification profile indicative of good reaction kinetics (exponential, linear and plateau phases). Strong sigmoid shapes are indicative to good amplification while flattened curves indicate that efficiency was less than ideal during the exponential phase. This may be indicative of inhibition or non-specific amplification. All 27 primer pairs displayed good sigmoid shapes indicative of good reaction kinetics.

Amplification specificity was assessed using both melt peak and MetaPhor gel electrophoresis. Melt peak analysis is readily facilitated by RT-PCR instrumentation, however it cannot differentiate between alternate amplicons with analogous melting temperatures, as such-high resolution gel electrophoresis was also used to identify mispriming (primer dimer or multiplex amplicon formation). Single melt peaks with relatively flat adjacent baselines indicative of target specific amplification with minimal primer-dimer formation were identified for all primer pairs bar IGF1_9 which displayed a two peak profile (T_Ms of 84.5 and 87.5°C). Melt curve analysis did not display any anomalies for IGF1_10 or BP3_14; however gel electrophoresis displayed triplet and double band formation with products of (246 / 207 / 199bp and 289 / 269bp respectively) as before.

Once amplification specificity was established, relative amplicon yields were assessed to determine optimal experimental T_A for all high-specificity primer pairs. The T_A at which highest single amplicon yields were derived was designated as the optimal T_A. Small primer-dimers and non-specific artifacts (such as those observed for BP3_1 at the lower

59.9°C annealing temperature) were considered when comparing amplification yields to avoid misappropriation of optimal T_A. The Panjkovich method displayed good predictive performance with 71% or primer pairs functioning optimally at the predicted T_A (~65°C) despite the relatively long primer lengths used [141]. A slight tendency towards underestimation of T_A may be noted; with 25% of primer pairs (including IGF1_5, IGF1_12, BP1_3/4, BP1_5, BP3_3, BP3_7) functioning optimally at T_A = 67.5°C. Just one primer pair (BP3_10/11) displayed optimal performance at T_A = 62.2°C. Relative yield analysis across the 8.9°C T_A spectrum tested indicate that all high-specificity primers (excluding IGF1_9, IGF1_10 and BP3_14) function well at 65.0°C, therefore strong multiplex amplification at this set annealing temperature should be possible.

Tolerance of primer pairs to sub-optimal conditions is known to suffer upon multiplex inclusion, as such the extent of primer pair robustness was also assessed by comparative analysis of yields derived at alternate annealing temperatures. Larger yield differences between products amplified at alternate T_As are indicative of primer pairs that display poorer performance in association with deviation from optimal T_A. High-specificity primer pairs were found to be quite robust to T_A adjustment with mean yield losses of 32.4% across this 8.9°C range. Primer pairs BP3_3, BP3_4 and BP1_3/4 displayed maximum yield losses of 64.8%, 55.9% and 53.4% respectively and were found to be least robust to deviations of this nature, while IGF1_1 and IGF1_8 are most robust to this kind of adjustment displaying just 14.5% loss across the T_A range tested. The reasons for deviation of this nature is unclear however the poorest performing pair BP3_3, contains two and three sequential adenine residues adjacent to a single cytosine clamp for forward

and reverse primers respectively. These weak adjacent sequences may have acted to reduce annealing efficiency and impact subsequent yield production.

4.4.2.2 Efficiency and dynamic range

Primer pair amplification efficiency and dynamic range has been shown to diminish upon inclusion in a multiplex configuration. PCR amplification efficiency reflects primer pair performance and is indicative of stable high-specificity yield production, as such determination of this feature is important as relatively small differences in efficiency can significant effect yield [221].

Efficiency is determined empirically and may be derived using sigmoid / logistic curve-fitting models or standard curve determination. Curve-fitting models work by fitting the experimentally derived amplification plots to a theoretical optimal sigmoid curve, while standard curve methods require amplification of multiple serial dilutions for construction of standard curves from which slope and efficiency are derived. Curve-fitting models offer an advantage in terms of the lower number of samples required to estimate efficiency, however standard curve determination allows efficiency to be determined across a broad dynamic range of template concentrations (up to 5 orders of magnitude for genomic DNA). Dynamic range is an additional measure of primer pair performance and reflects the ability of primer pairs to function optimally across a wide range of target concentrations. Its determination is most often employed for assessment of primers used in expression analysis, where a range of target concentrations must be accurately amplifiable. Although only a single target concentration is to be used during our analysis, multiplex inclusion is known to reduce primer pair capacity and as such primer pairs

displaying poor singleplex dynamic range may be rendered non-functioning by inclusion in higher dimensionality formats. All primer pairs were therefore tested using the standard curve method for assessment of efficiency across a dynamic range four orders of magnitude. A six point standard curve spanning a range of target concentrations from $6.16x10^{1}$-$6.16x10^{-2}$fMole of genomic DNA was constructed for all primer pairs. Amplification curves, melt peaks and electrophoresis were used to assess primer pair specificity and kinetic adherence as before. High-resolution electrophoretic analysis showed all non-repeat region pairs to be highly specific, producing a single target band of anticipated size. Repeat region pairs IGF1_9, IGF1_10 and BP3_14 again displayed multiple band formation.

Efficiency was calculated using Bio-Rad CFX method which implements the Pfaffl and Vandesompele formula with percentage conversion [222,223]. Using this method, efficiency should range between 90-110% for primer pairs with good functionality. The slope of the standard curve is directly related to the average efficiency of amplification and should be between -3.6 and -3.1 while r^2 (correlation coefficient) indicative of the quality of the fit of the data points plotted to the standard curve and should be >0.95. Primer pairs with scores exceeding set limits in any categories are determined to display "poor" amplification efficiency. A total of 24 high-specificity primer pairs displayed good amplification efficiency across this dynamic range with mean efficiency, slope and correlation coefficients of 101.90, -3.27 and 0.988 respectively.

Primer pair IGF1_8 displayed poor efficiency of 114.3% and a suboptimal slope of -3.021. At 72bp, IGF1_8 is designed to produce the shortest amplicon in the set, as such

its requirement in terms of optimal extension times are below that of other pairs, a factor which may have caused the poor performance observed in this instance. This result may highlight potential IGF multiplex set incompatibility with regard to use of single extension times for amplification of alternately sized products. Primer pair IGF1_9 also displayed sub-optimal performance using this method with poor efficiency, slope and correlation coefficients of 153.9%, -2.478 and 0.775 respectively. This poor performance is not unexpected given its poor performance in previous assays and the nature of non-specific amplification formerly observed.

Singleplex optimisation protocols are useful to help determine primer pair weaknesses (poor specificity, efficiency, alternate optimal T_As etc) however PCR protocols and reaction mixes are altered for multiplex inclusion and as such these changes may ameliorate or inhibit performance of primer pairs in a manner not anticipated by previous singleplex studies. Analysis of this type is performed to guide identification of poorly performing pairs, but determinations may not fully translate in subsequent multiplex reactions and optimisation of higher dimensionality formats may be required.

4.4.3 Multiplex PCR

Once extensive singleplex amplification procedures had been performed and evaluated, end-point multiplex amplification was initiated [224,225]. Smaller dimensionality 5-plexes were initially amplified using the Qiagen multiplex protocol as described and analysed using high resolution MetaPhor gel electrophoresis (see electronic appendix Figure 3 and Figure 4). Amplification of these multiplexes was relatively straightforward with common reaction mixes, thermocycle profiles and equimolar primer concentrations

found to produce relatively strong yields for all pairs, without primer dimer or non-specific artifact formation. Resolution and sensitivity was poor using MetaPhor / ethidium bromide staining however and as such a higher resolution / sensitivity alternative was sought. The Experion automated electrophoresis system was used to achieve these ends. Using company specified guideline resolution of up to 5bp should be achievable within the target range. Amendment of this protocol by cooling reagents and chip to ~1°C prior to analysis was found to increase resolution allowing separation of fragments differing by as little as 4bp (BP3_7 and BP1_3/4, 152 and 156bp respectively) in the <160bp range.

Once the maximum resolution capacity was established, the highest dimensionality single-tube multiplex sets were constructed. Four multiplexes containing 14, six, five, and two target regions, covering 27 bins across candidate genes IGF1_1, IGFBP1 and IGFBP3 were assembled. Downstream suspension array hybridisation are effected in part by target concentration; as such optimisations were performed to achieve relatively equimolar amplification of targets with all amplicon yields optimised to displaying less than 50% yield variance between the largest and smallest yielding primer pair [153]. Reaction components, relative primer pair concentrations and thermal cycle profiles were adjusted to compensate for bias based on the individual needs of each multiplex and previous singleplex investigations.

4.4.3.1 2-plex Amplification

Previous singleplex analysis determined IGF1_9 to be susceptible to non-specific artifact formation; as such IGF1_9 was purposely assembled into the lowest dimensionality 2-

plex construct to minimise further complications imposed by more complex mixtures. An Ampli*Taq* gold derived protocol (previously shown to facilitate target specific amplification of IGF1_9) was used in place of the Qiagen multiplex mix which had been shown to induce nonspecific amplification of this target. Both IGF1_9 and IGF1_11 were found to produce 15 and 14 ng/µl of target respectively using singleplex Ampli*Taq* Gold amplification (and end-point quantitation). Using the multiplex method as described, IGF1_9 and IGF1_11 produced relatively equimolar target specific amplicons of 10.48 and 12.4ng/µl with minimal primer-dimer formation. The low dimensionality of this multiplex mix is likely to have contributed to the relatively straightforward nature of amplification observed; relative primer pair concentration adjustment was found to be unnecessary in this instance.

4.4.3.2 5-plex /6-plex Amplification

Using Qiagen multiplex PCR reaction protocol and thermocycle profile as described; target specific amplification was achieved for both 5-plex and 6-plex reactions.

Application of 5-plex mix containing equimolar primer pair concentrations of BP1_3/4, BP3_10/11, IGF1_6, IGF1_2 and BP3_1 resulted in high specificity amplification of targets of anticipated size; 156, 168, 174, 243 and 250bp (+/- 17bp). End-point efficiency divergence was high however with BP3_10/11 and BP3_1 displaying significantly higher relative efficiencies of 1 and 0.519 compared to those derived for BP1_3/4, IGF1_6 and IGF1_2 (0.296, 0.235 and 0.230 respectively). RT-PCR singleplex yield divergence for all five primer pairs was determined to be 22%, this was increased to 77% upon multiplex inclusion. Optimisation via primer pair concentration adjustment was therefore performed to compensate for alternate efficiencies within this format. Using this approach primer

concentrations of 0.3-1.0µM were used to produce five target amplicons with a narrow 28% yield variance ranging from 8.82-12.31 ng/µl for IGF1_2 and BP3_1 respectively.

6-plex primer pair amplification (BP3_13, BP1_5, IGF1_5, BP3_12, IGF1_7 and IGF1_3) also resulted in target specific amplification, with six amplicons of anticipated size 154, 168, 175, 205, 234 and 248bp (+/-13bp) produced. Again multiplex amplified targets displayed higher relative yield divergence (~73.8%) compared to those derived using singleplex RT-PCR amplification (31%). Primer pairs BP3_13 and IGF1_5 displayed high relative end-point efficiencies of 1.0 and 0.824 compared to an average efficiency of 0.349 for all other pairs in the set. Again relative primer pair concentration adjustment was used to compensate for divergent yields. Using final primer pair concentrations 0.5-2.0µM; six target amplicons ranging from 7.26-12.45ng/µl for IGF1_7 and BP3_12 respectively were determined equating to a yield variance of 41.69%.

4.4.3.3 14Mpx Amplification

In accordance with Rachlin *et al*., assertion that "achieving broad SNP coverage rapidly transitions from being very easy to very hard as the target multiplexing level (# of primer pairs per tube) increases", equimolar amplification of the 14 targets included in the largest multiplex construct was more challenging than previous smaller amplifications [226]. Total primer and reagent consumable concentrations, kinetic considerations and innate primer features all contributed to make multiplex optimisation significantly more challenging in this instance.

Amplicon length compatibility issues were raised by this assembly, containing the longest possible range of targets spanning 428bp from 72-500bp (for IGF1_8 and

IGF1_12 respectively). As before, equimolar primer pair concentrations were initially used for multiplex amplification. Thermocycle annealing and extension times of 1 min 30 seconds (as recommended by Qiagen multiplex PCR protocol) were initially used for 14-plex amplification. This facilitated amplification of 13 targets, however amplification of the largest 500bp (IGF1_12) amplicon was unsuccessful. Annealing time was duly increased in 30 second increments to a maximum of 3 min 30 seconds annealing and the effect on amplification noted. Longer, 3 min annealing times produced the largest IGF1_12 yield, while retaining all other target amplicons. Annealing times exceeding this threshold did not act to improve amplification and overall yields were reduced.

Polymerase displays strong activity in the range of 65 to 78°C; as such the high 65°C annealing temperature used during this protocol should facilitate primer elongation during both extension and annealing phases; mirroring the reaction profile of two-step PCR. DNA polymerase displays reduced activity with increasing temperature exposure; as such increasing the total combined duration of annealing and extension past 4min30sec was not feasible. Due to increased thermal conductivity rates, it was postulated that "true" segment times may be increased by reducing reaction volumes. As such, a range of reaction volumes from 25-100ul were used for 14-plex amplification. 100µl reaction volumes resulted in preferential amplification of smaller targets, while use of 25µl volumes induced preferential amplification of larger targets (in particular IGF1_12). However, while use of smaller volumes aided IGF1_12 amplification, it also resulted in failure of the smallest IGF1_8 (72bp) target amplicon and poor overall yields. Although not equimolar; 50µl reaction volumes produced successful amplification of all 14 targets with minimal non-specific amplification. As such, a 3 min / 1 min 30 sec annealing /

extension profile in conjunction with a total reaction volume of 50µl was selected for further optimisation.

Using equimolar primer pair concentrations; target-specific amplification of all 14 targets was performed with yields and relative efficiencies determined. Relative efficiency within the 14-plex ranged from 1-0.296 for BP1_2 and IGF1_12 respectively with a mean relative efficiency of 0.607. A correlation appears to exist between amplicon length and efficiency with both largest IGF1_12 and smallest IGF1_8 primer pairs performing quite poorly relative to those in the optimal 200bp range (100-300bp). Aside from amplicon length, repeat region inclusion and GC clamp exclusion (previously discussed), all other design features thought to affect amplification performance were strongly adhered to within narrow optimal boundaries (GC% and distribution, intra and inter primer Tm etc). Accordingly, no correlation between any of these features and relative primer pair efficiency was observed. Potentially problematic AutoDimer-predicted heterodimer formation between IGF1_1 and BP3_3 forward primers also appeared to have little appreciable effect on amplification with strong relative yields of 0.716 and 0.677 determined for these pairs within the 14-plex structure.

Although primer pair concentration adjustment was performed as before, an unforeseen effect was noted; use of increased total primer concentration was found to induce preferential amplification of smaller targets at the expense of larger targets (especially 500bp IGF1_12 amplicon). The effect of total primer concentration on amplification efficiency using this 14-plex structure was therefore investigated. Using a range of total primer concentrations target-specific amplification was achieved. Skewed large and small

target amplification profiles abound however, at both ends of the 1µM and 3µM spectrum respectively. Non-specific amplification was also seen to have occurred with non-specific amplicons in the range of 22-65bp and 710-1121bp evident for 1.0µM and 3µM reaction mixes respectively (Figure 63).

An optimal total primer pair concentration of 2.0µM was identified and relative concentration adjustment performed to this specification. Using final primer pair concentrations of between 0.04-0.15µM (BP3_10/11 and IGF1_12) equimolar amplification of all 14 targets was achieved with amplicon yields ranging 47.09% from 4.36 - 8.24ng/µl for IGF1_13 and IGF1_8 respectively.

Following the removal of non-specific IGF1_10 and BP3_14 primer pairs, a four-test structure, simultaneously amplifying 14, 6, 5 and 2 target loci across 99kb was achieved using multiplex PCR. To the best of our knowledge; the larger 14-plex IGF construct developed during the course of this work is among the highest dimensionality multiplex assembled to date for targeted gene-specific disease association and with 14 primer pairs spanning ~99kb (1 assay per ~7kb). This is just below the largest commercially available high-density gene specific multiplex test developed to date which screens 31 cystic fibrosis mutations across 188kb (1 assay per ~6kb).

The work in this chapter aimed to meet the objective "*to perform extensive PCR optimisation for the construction of a number of robust, well characterised, high dimensionality multiplex PCR sets*" as described. This objective was met. All primer pairs (designed in chapter 3) were assessed in terms of their singleplex PCR specificity, efficiency and tolerance and this data used to direct construction of a 14-plex, 6-plex, 5-plex and 2-plex sets which display high-specificity amplification of target loci. The following chapter aims to demonstrate the applicability of this type of multiplex amplicon target to suspension array genotyping using a model IGF probe pair designed in chapter 3.

Chapter 5: Suspension Array Genotyping

5.1 Introduction

5.1.1 Overview

The fundamentals of suspension array technology (SAT) in association with a number of the advantages of this approach relative to other SNP genotyping formats has previously been outlined in chapter 3. The following section describes experimental design considerations and divergent suspension array-based options which may be employed to meet the increased multiplex, high-throughput, robust requirements of suspension array disease association experiments. The applicability of this technique to IGF1 multiplex set probe pair pIGF1_1C/pIGF1_1T is also demonstrated in a proof of concept suspension array allele discrimination study.

5.1.2 Encoded multiplex microspheres

Functionalised microspheres are the solid support probe carrier of choice for suspension array-based genotyping assays [13]. A number of commercially available sets have been developed and are generally composed of inert material such as Poly(methyl methacrylate) (PMMA), polystyrene or silica, with sizes ranging from of 0.5-10µM [227-230]. They are manufactured to display low variation in terms of microsphere size, encoding and stability distributions allowing simplified population discrimination using a host of fluorescent reporters and detection platforms. Microspheres are generally functionalised using amination, carboxylation, avidination or biotinylation. Functional group densities and intergroup distances are set to allow for a host of molecule types including antibodies, enzymes and oligos permitting the application to genomics and proteomics [227-231].

5.1.3 Microsphere encoding

A wide array of encoding schemes have been reported in the that can be broadly categorised into non-optical and optical (fluorophore or nanocrystal) approaches. Non-optical encoding schemes exploit physical dissimilarities for classification of alternate microsphere populations. Particle size, surface-enhanced resonance Ramon spectra effect (SERRS), radio frequency tagging and fluorophore or quantum dot encoding have all been utilised. Particle size was initially used by Horan *et al*., to discriminate between alternate beads with as little as 0.1µm size differences, however the scope of such sets is limited and so alternate approaches were investigated. Jin *et al*., exploited the increased sensitivity and narrow spectral bandwidths afforded by SERRS to achieve high sensitivity microsphere encoding with increased multiplex dimensionality [12,232]. The SERRS effect occurs when molecules adsorbed on rough metal surfaces (silver or gold) display Raman scattering and the incident light is resonant with both the molecule and plasmon of the metal. This results in increased Raman scattering (by up to 10^{15}) allowing femtomolar to attomolar target detection, consequently Raman peaks with narrow bandwidths of ~1 nm facilitate high dimensionality multiplexing [232-234]. However, while SERRS SAT can increase sensitivity and multiplex capacity, decoding Raman spectra is complex; a feature likely to discourage widespread uptake of this approach [235].

Radio frequency tagged microchips have also been developed with good effect. These microchips are fitted with transponders which transmit a distinct radio frequency which may be used to determine microsphere identities. Moran *et al*., used capillary electrophoresis and laser activated signal transmission for the discovery of tripeptide-substituted cinnamic acid inhibitors of the protein tyrosine phosphatase however while

this approach offers a large capacity for multiplexing; transponders are relatively large, a feature disadvantageous for oligonucleotide analysis [236].

Optical encoding using fluorophores or quantum dot doping is the most popular encoding mechanism employed for microsphere identification. This is due to the ease of identification and high throughput analysis facilitated by use of widely available flow cytometers and bulk encoding strategies which allow large quantities of relatively inexpensive microspheres to be produced with high uniformity and minimal size, intensity and granularity distributions [13,237-240]. Using this approach distinct spectral barcodes are produced by combining a number of fluorochromes with alternate emission wavelength at fixed intensities (see Figure 65).

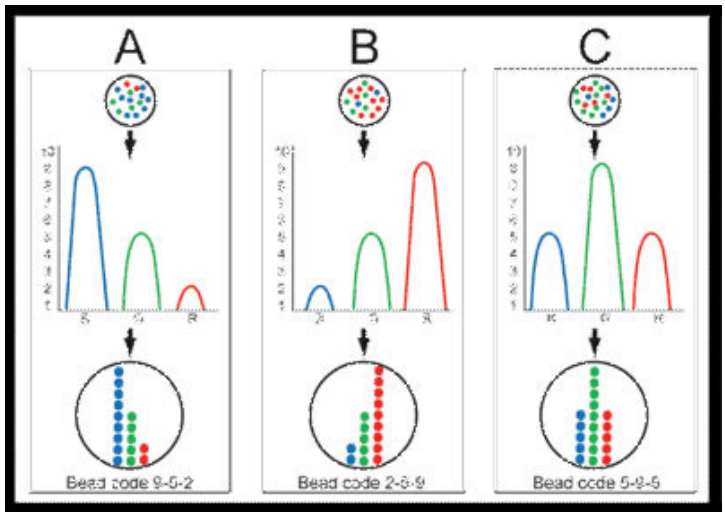


Figure 65: Microsphere Optical Encoding. Microspheres are doped using multiple fluorophores with alternate emission wavelengths at fixed intensity ratios to produce an array of distinct spectral barcodes for alternate microspheres within a combinatorial library [241].

Using the equation below, 1,900, 000 spectral codes may be produced using 6 emission wavelengths and 11 intensity ratios. In practice however decoding such combinatorial sets would be challenging and a number of other experimental limitations with regard to fluorochrome excitation and emission spectrum also restrict set sizes to lower more manageable dimensions.

$$C = Nm - 1$$

Where

- C = number of unique codes possible,
- N = number of intensity levels and
- m = number of emission colours

Organic dyes have been predominantly used for optical encoding and are available in a number of commercial kits which allow simultaneous analysis of up to 100 microspheres [229,231,242,243]. Organic dyes have narrow excitation and broad emission spectra which display red tailing and as such combinatorial sets of this nature are limited in terms of the number of fluorophores which can be included and differentiated during analysis. Development of semiconductor nanocrystals or quantum dots have acted to loosen combinatorial constrains allowing the scope of higher dimensionality sets to be more broadly investigated [244].

Quantum dots are nanometer scale moieties (2-10nm) composed of semiconductor materials including Cadmium Selenide and Zinc Sulfide whose physical dimensions are lower than that of the Bohr radius[11]. When photons of light strike semiconductor materials electrons are excited and are elevated to higher energy levels, upon their return to ground state they generally release a photon of light characteristic of the semiconductor material, however when particle size is lower than that of the Bohr radius, band-gaps between ground and excited states become size dependant and as such semiconductors display emission spectra whose wavelengths are directly proportional to particle size. This is termed the quantum confinement effect [238,245].

[11] The Bohr radius is the mean radius of the orbit of an electron around the nucleus of a hydrogen atom at its ground state (lowest-energy level). Bohr radius represents the smallest mean radius normally attainable by a neutral atom.

This effect can be exploited to produce a wide range of quantum dots from a single semiconductor material with differing emission spectra. Quantum dots may be assembled into bead-based formats containing differently sized nanoparticles with a range of emission maxima at fixed ratios; producing beads with distinct spectral barcodes. Quantum dots display a number of advantages over biological fluorophores which increased ease of handling, minimal photobleaching, large molar extinction coefficients, high quantum yields, broad excitation wavelengths and narrow emission spectra (25-40nm FWHM) as well as their capacity for high-combinatorial set construction [240]. As such, quantum dot-encoded microspheres offer a facility for high-dimensionality suspension array applications and work has begun with regard to commercial development of sets of this nature [246,247]

5.1.4 Detection instrumentation

A range of existing and novel detection platforms have been applied to suspension array analysis. These platforms should ideally display high sensitivity, accuracy, precision, throughput and multiplex capacity at minimal cost. The following section describes two approaches to suspension array detection which utilise static and flow based platforms [248-250]. The mosaic system is a dedicated static platform developed specifically to facilitate robust quantitative oligonucleotide analysis using a static platform and quantum dot-encoded microspheres (QDEM). Using a single laser for excitation of multiple QDEM encoded microsphere species in conjunction with a coupled device (CCD) image detector, multiplex analysis of sedimented microspheres may be performed [251]. This system offers a number of advantages including use of low-cost CCD detectors and a high level of automation however this product has now been discontinued and a

replacement static system has yet to emerge. Additionally concerns remain regarding coagulation of microspheres and overlapping species can contribute to the difficulty associated with data interpretation using this approach [249,250].

Flow based platforms offer the opportunity for single microsphere analysis and the high accuracy which this approach affords. The Luminex 100 platform is a dedicated dual-laser flow-based system analogous to flow cytometry specifically designed for Luminex 2-colour microspheres analysis. The system has been applied to multiple applications including nucleic acid profiling, immunoassays and cytokine analysis [252-259]. The specific application of this design means that assay set-up times are minimised and compensation requirements are reduced, however the system is less flexible than standard flow cytometers and Tsuchihashi *et al.,* suggest that due to its use of biological fluorophores its capacity is limited beyond its current capacity [247,260].

Flow cytometry (FC) is a well established technique which facilitates simultaneous multiparametric analysis of microsphere or cell characteristics (physical or chemical) using flow-facilitated manipulation of target moieties through an optical analysis and detection system. Flow cytometers are widely available in hospitals, universities and core laboratories and have been developed to include multiple lasers and PMTs which can detect up to 14 emission wavelengths, however these more complex FC systems are incredibly expensive and the bulk of current flow cytometry-based applications are performed on single/dual-laser, 4 PMT detector FCs [261]. The unique properties afforded by quantum dot-encoded microspheres (QDEMs), including single wavelength excitation and narrow emission spectra, allow these basic cytometers to be used for more complex genotyping and proteomic assays. They provide an easily accessible high-

throughput analysis platform which is flexible, sensitive, and accurate allowing a high degree of multiplexing. The following section describes in more detail the mechanism by which FC instrumentation functions and its putative capacity with regard to SAT sample analysis.

5.1.5 Flow Cytometry

Flow cytometry is a powerful technique for the characterisation of multiple parameters from individual cells or microspheres ranging from 0.5-100μM[12] within heterogeneous population samples at analysis rates of up to 100,000 sec^{-1} [13]. During sample analysis particles are injected into the centre of a fast flowing sheath fluid stream where, due to the principles of laminar flow the sheath and sample fluid do not mix and pressure differences may be used to hydrodynamically focus the sample particles into a thin corridor of single concurrent particles which are analysed at the interrogation point. The interrogation point consists of a clear section of capillary tubing which is traversed by a laser and flanked on multiples sides by detectors and a series of optical filters.

When a particle passes the detection window it intersects the laser causing light to be scattered, absorbed and emitted in a manner directed by the innate particle profile. Side scatter detected at 90° to the incident of light is a measure of particle granularity or structural complexity, while light scattered in a predominantly forward direction near the angle of incidence is a function of cell diameter. Scattered light is quantified by a detector that converts intensity into a voltage pulse. The scatter profile and therefore voltage pulse is proportional to the particle size or complexity, with larger and more complex particles producing larger signals on forward and side scatter detectors respectively. Histograms

[12] Bacteria and blastocysts respectively

may be used to plot forward and side scatter signals to determine sample population size and complexity distributions determination. Use of 2D scatter plots (SS versus FS) can allow increased resolution and differentiation of subpopulations within the sample, while additional fluorescent measurements first exploited by Wolfgang Göhde in 1968 may be used to determine a wide range of non-physical particle characterises through fluorescent labelling [262]. Fluorescent emission occurs when molecular absorption of photons cause electrons to become excited and move to higher energy states, when electrons return to ground state they release photons of light at longer fluorescent wavelengths. By coupling fluorescent moieties to specific cellular or microsphere targets, characteristic determination regarding the target of interest may be made [263,264].

Fluorescent photons are directed along the same path as the side scatter signal and manipulated through a series of filters and mirrors to allow separation of multiple fluorescent wavelengths into ranges detectable by dedicated photomultiplier tube detectors. Fluorescent signals are detected in a manner analogous to scatter data; with fluorescent light intensity converted to voltage signals in a manner indicative of intensity. Multiple fluorophores may be used to determine alternate characteristics for a single particle. The number of characterises determinable is dictated by the number and scope of PMTs, lasers and fluorophore excitation and emission spectra. Use of quantum dots in place of biological fluorophores can act to increase this capacity allowing improved multiplexing and data acquisition from single particles [263,264].

The work described in this chapter sets out to meet the final project objective "*To demonstrate the applicability of multiplex PCR to suspension array facilitated allele discrimination using a model probe pair and validate allele designation via dot blotting*". During the course of this work, the potential application of quantum dot-encoded

methacrylate microspheres to suspension array techniques were assessed. Suspension array coupling and hybridisation approaches were evaluated in a proof of concept study using model probe pair pIGF1_1C/pIGF1_1T in association with singleplex IGF1_1 and multiplex 14-plx targets and assay boundaries determined. The following section describes experimental protocols implemented throughout this SAT study.

5.2 Methodology

All FCM was performed using a Beckman Coulter EPICS® XL flow cytometer with 488nm air-cooled laser, four-colour fluorescent filter set-up (252BP, 575BP, 625BP and 675BP) and digital signal processing [265]. Optical alignment and FC fluidic performance was assessed using SPHERO™ Ultra Rainbow Calibration Particles and instrument maintenance and quality control performed in accordance with Beckman Coulter recommendations [266,267]. Data analysis was performed using De Novo FCS Express software [268]. All reagents used during coupling and optimisation protocols were purchased from Sigma Aldrich unless otherwise stated.

5.2.1 Microsphere Characterisation and Selection

The PlxBead quantum dot-encoded microsphere solid support library was purchased from Crystalplex (Crystalplex Corp., Pittsburgh, PA). This combinatorial assembly is composed of 5µM mesoporous methacrylate microspheres with carboxyl functionalisation. Each of the 16 microsphere species has a distinct spectral address, encoded using a maximum of four emission wavelengths (525, 575, 620 and 675nm) with binary on/off intensities.

Microsphere counts were performed on untreated microsphere samples using haemocytometer counting and flow cytometry to determine microsphere populations

(particles / µl) [269,270]. Microsphere handling was performed in accordance with manufacturer instructions.

In order to assess the response of microspheres to assay-relevant conditions microspheres were incubated with a number of solutions including H_2O, TE, Phosphate-buffered saline (PBS), Tetramethyl Ammonium Chloride (TMAC) spanning a range of salt (0-4M) and pH conditions (pH4 -10) over discrete time periods (15 min – 48 hours). The effect these treatments on microsphere populations were assessed using haemocytometer and flow cytometry counting as before.

A second set of distinct microspheres were purchased from Spherotech (Spherotech, Inc., Lake Forest, IL) and population counts under the range of conditions previously specified repeated. Unlike previous microspheres these particles are fluorophore-encoded and composed of polystyrene. This 2-plex set consists of carboxyl functionalised microsphere codes, 5.0-5.9 µm in size with single emission wavelengths at 530nm and 635nm for microspheres CFP-5052-2 (Carboxyl Fluorescent Yellow Particles) and CFP-5067-2 (Carboxyl Blue Particle Array Chemistry Development Particles). These microspheres are termed S1000 and S0001 respectively throughout the thesis for ease of recognition.

5.2.2 Quantitative Flow Cytometry

The applicability of quantitative flow cytometry to EPICS® XL (Beckman Coulter, Inc., Fullerton, CA) analysis was assessed using Spherotech UltraRainbow Calibration Particle URFP-38-2K. Flow cytometry voltage and gain values (Table 33) were adjusted to allow particles to be observed across three FC channels (470, 575 and 675nm) with five microsphere intensities spanning four orders of magnitude as recommended. The population was gated on forward and side scatter and a count threshold of 10,000 applied.

For each PMT detector; histograms were used to determine channel (peak) numbers for all five microsphere intensity levels, channel numbers were plotted against Spherotech determined relative molecules of equivalent fluorescence (MEF) and calibration curves potted. Correlation coefficients and curve shapes were used to assess the suitability of this technique to our instrumentation (i.e. the EPICS® XL).

Table 33: EPICS® XL Quantitative Flow Cytometry Settings. PMT voltage and gains are specified for forward, side and auxiliary channels as well as fluorescence channels F1, FL2, FL3 and FL4. Populations were gated on forward and side scatter and compensation was not applied.

	Voltage	Gain	
FS	432	2	4.59
SS	39	10	11.17
FL1	819	1	-
FL2	820	1	-
FL3	795	1	-
FL4	1021	1	-
AUX	500	10	25.00
Compensation	None		
Gate	FS/SS		

5.2.3 Microsphere Coupling

Oligonucleotide probe immobilisation was facilitated by heterobifunctional EDC (1-Ethyl-3-[3-dimethylaminopropyl]) coupling between carboxylated microspheres and aminated oligonucleotide probes. A synthetic PE-Cy5 labelled aminated-Poly(dA) reporter probe with FL4 (670nm) emission facilitated coupling optimisation. 2µl of Spherotech S0001 0.5% w/v partiles (2.9×10^4 microspheres) were washed (x2) with 0.1mol/L 2-(N-morpholino) ethanesulfonic acid (MES) pH 4.5. Microspheres were resuspended in 20µl MES and 150µl of fresh EDC (250g/L) added. The solution was vortexed for 1min 45sec (Fisons Whirlimixer WM250-SC) to activate microsphere

carboxyl groups and further MES (×2) wash steps performed [153,231,271,272]. Microspheres were resuspended in 40µl MES and 1.00pmol heat-denatured/ snap-cooled (95°C 5min/ 0°C 5min,), probe added to the mix. The solution was vortexed, sonicated (1min Sonicor SC-121) and incubated at room temperature with rotation (instrument and setting) for 2 hours. Following incubation coupled microspheres were washed in weak non-ionic detergent 0.02% Tween, 0.1% SDS to remove non-specifically bound probe species [273]. Microspheres were resuspended in 600µl PBS for immediate FC analysis or stored in 0.1M imidazole (pH7.0). A number of optimisations were performed during the course of protocol development including optimisation of probe concentration (0.25-2.00pmol/µl) and incubation duration (15min – 3hour).

5.2.4 Microsphere-Target Hybridisation

Following coupling of S1000, S0001 microspheres and target specific probes pIGF1_1T and pIGF1_1C, hybridisation was perfumed between microsphere-probe species and PCR targets. 49.5µl of 75mM Tris-HCL (pH8.0) buffered hybridisation solution containing 4.5M TMAC salt, 0.15% sarkosyl and 6mM EDTA as described by Dunbar *et al*., was added to coupled microspheres (~20,000) [274]. PCR product was added to sample tubes to produce final IGF1_1 target concentrations of between 5-50fmol. 1x TE buffer was added to make a final volume of 75µl. Samples were denatured (95°C, 10min, *Techne TC-512*) and hybridisation performed at 60°C for 2 hours (*Hybaid* Maxi 14). Hybridisation duration (15 min -2 hrs) and temperature (40-70°C) were optimised for maximum allele discrimination and optimal concentration ranges for singleplex and 14-plx targets determined. Following hybridisation; samples were washed (× 2) in 1× hyrbidisation buffer and resuspended to a final concentration of 6mg/µl streptavidin PE-Cy5. Streptavidin-biotin coupling was performed at 37°C for 30 minutes and samples

subsequently washed in 0.02% Tween/0.1% SDS to remove non-specifically bound molecules. Microspheres were resuspended in 1,200µl PBS for FC analysis.

5.2.5 FC Sample Analysis

Geometric mean fluorescence intensity (MFI) determination was used to assess coupling efficiency for each microsphere population as follows: Normalised geometric mean (nMFI) = Processed microsphere MFI – unprocessed microsphere MFI.

EPICS® XL FC analysis for both coupling and hybridisation reactions were performed using a low flow rate, total sample volume of 600-1,200µl, FS / SS gating and an event threshold of 5,000 / 10,000 particles for singleplex and duplex bead code samples respectively. Voltages and gains as described (Table 34) were applied for analysis of all coupled and hybridised samples.

Table 34: EPICS® XL Sample Analysis Flow Cytometry Settings. PMT voltage and gains are specified for forward, side and auxiliary channels as well as fluorescence channels F1, FL2, FL3 and FL4. Populations were gated on forward and side scatter and compensation was not applied.

	Voltage	**Gain**	
FS	432	2	4.59
SS	39	10	11.17
FL1	386	1	-
FL2	409	1	-
FL3	410	1	-
FL4	637	1	-
AUX	500	10	25.00
Compensation	None		
Gate	FS/SS		

5.2.6 Dot Blot

Dot blotting was performed using GE healthcare ECL Direct Nucleic Acid labelling and Detection System according to manufacturer's instructions. Singleplex IGF1_1 PCR product was chemically denatured using 1.5M NaCl, 0.5M NaOH solution and spotted onto Hybond-N^+ nitrocellulose membrane (25ng per spot). Spots were fixed using 30sec exposure to UV light (Herolab Clean Cab) and incubation with probe sequences IGF1_1C, IGF1_1T and IGF1_5G performed at 42°C ~6hrs (Hybaid Maxi 14). A range of primary stringency washes (0.4% sodium dodecyl sulfate with 0.2X, 0.3X, 0.4X and 0.5X SSC) were applied to alternate repeat membranes in order to facilitate allele discrimination via spot intensity quantitation.

5.3 Results

5.3.1 Microsphere Characterisation and Selection

Extensive bead loss was observed in initial suspension array-based experiments using methacrylate Crystalplex PlxBeads. It was hypothesised that high salt concentrations or pH adjustment may cause osmotic shock and degradation of PlxBead material or leaching of QDs. The effect of a number of relevant buffer solutions and pH environments was therefore assessed to determine whether these induced PlxBead loss (Figure 66, Figure 67 and Figure 68).

Microspheres were determined to be unaffected by H_2O incubation. Some loss was observed using TE buffer incubation however this was manageable within the confines of SAT time requirements. Microsphere loss was critical using both PBS and TMAC (2.5M) solutions however. PBS buffer contains phosphate which was subsequently found to causing leeching of quantum dots from the polystyrene shell (Crystalplex Corp., personal correspondence). A high salt TMAC (~2.5M) solution was found to have the most negative effect on microsphere stability (Figure 66). It was hypothesised that osmotic shock may induce microsphere breakage and therefore the effect of increasing TMAC concentration 0.01-4.00M on bead population maintenance was tested following 15minutes incubation. A clear correlation between increasing TMAC concentration and reduced microsphere population was evident with 4M TMAC sample displaying a PlxBead population of just ~20% relative to the size of lowest 0.01M concentration tested (Figure 67). The effect of TMAC concentration (0.1-0.01M) over a range of pHs 4-10 was also assessed. Microsphere populations were found to be most stable at pH10 in low salt concentrations (0.01M) (Figure 68).

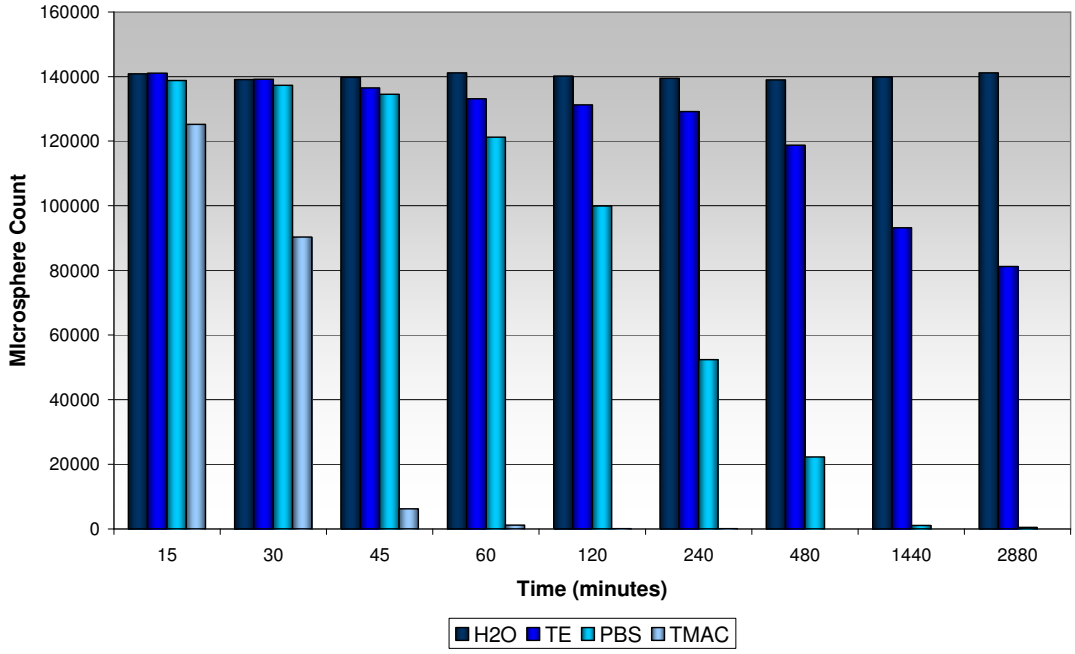


Figure 66: The effect buffer solution on PlxBead population yields was assessed at intervals over a 48hour period. Non-significant microsphere loss was determined for H2O incubation; loss was manageable using TE buffer within the confines of SAT time requirements however microsphere loss was critical using both PBS and hybridisation buffer. PBS buffer contains phosphate which causing leeching of quantum dots from the polystyrene shell (personal correspondence Crystalplex). The reason for hybridisation buffer incompatibility is unknown however this solution contains a high salt concentration (~2.5M) relative to alternate solutions tested, as such osmotic shock or salt sensitivity may effect bead disruption and breakage.

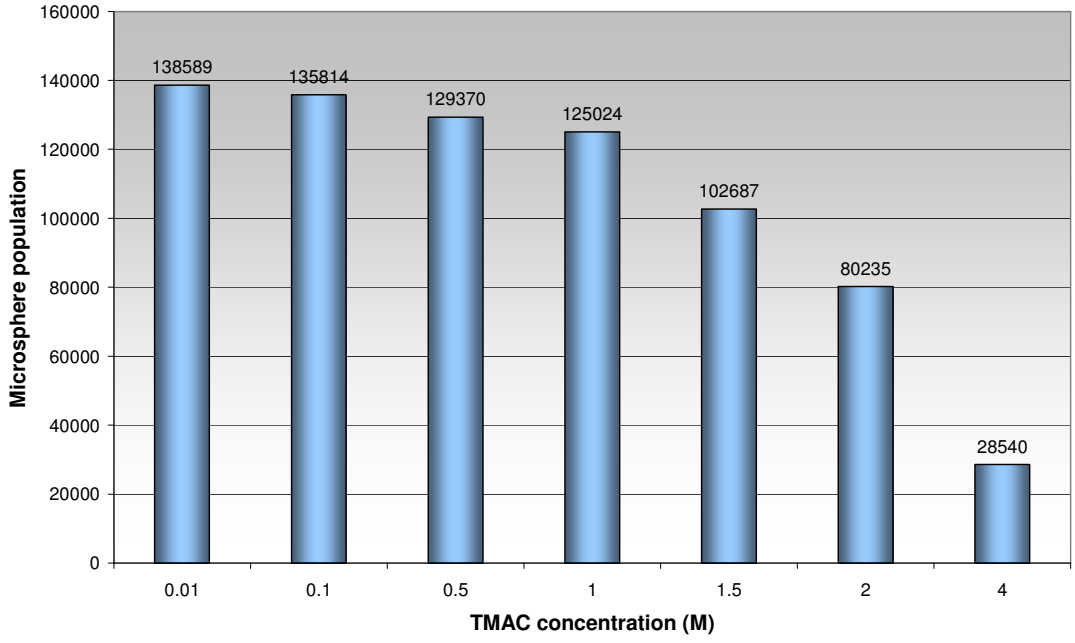


Figure 67: The effect of TMAC (tetra-methyl ammonium chloride) concentration on PlxBead population as determined following 15 minutes incubation and FC analysis. A strong correlation between increasing TMAC concentration and PlxBead loss may be seen to occur with the highest 4M TMAC sample displaying a PlxBead population of just ~20% relative to the size of lowest 0.01M concentration TMAC sample.

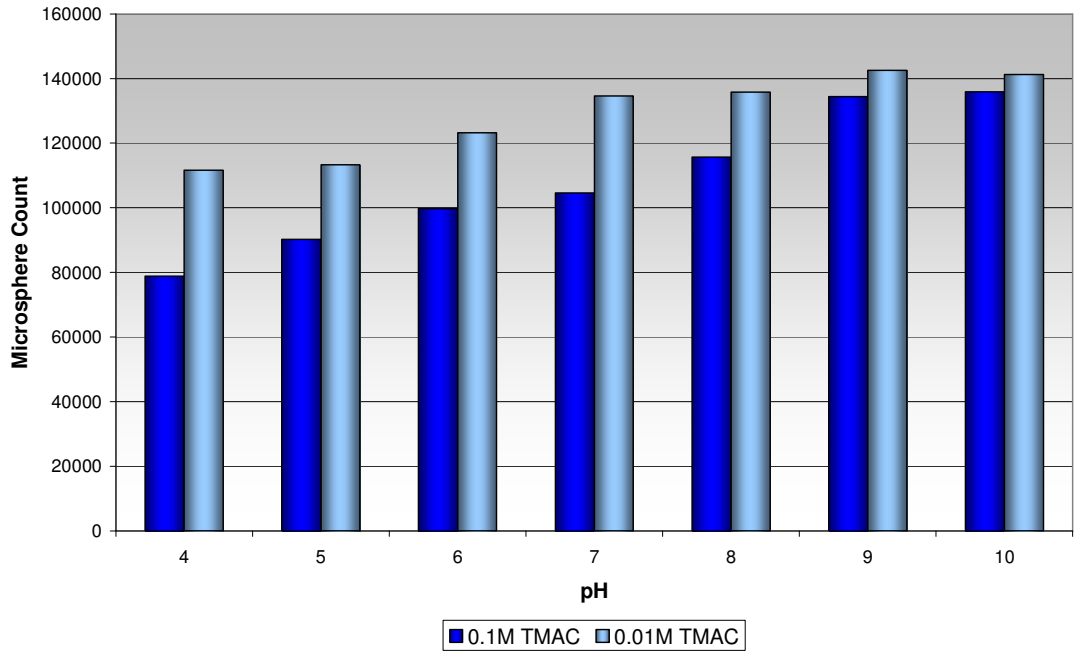


Figure 68: Effect of pH (4-10) on PlxBead population following 15min incubation in 0.1 and 0.01M TMAC solutions. PlxBeads appear most stable in high pH (~pH10) and lower 0.01M TMAC concentrations.

5.3.2 Quantitative Flow Cytometry

The applicability of quantitative flow cytometry to EPICS XL analysis was assessed using Spherotech UltraRainbow Calibration Particle URFP-38-2K. Flow cytometry volt and gain values were adjusted to allow particles to be observed across three FC channels (470, 575 and 675nm) with five microsphere intensities spanning four orders of magnitude as recommended. For each PMT detector; histograms were used to determine channel (peak) numbers for all five microsphere intensity levels (Figure 69 and Figure 70). Channel numbers were plotted against Spherotech determined relative MEFs (molecules of equivalent fluorescence) and calibration curves potted (Figure 71 and Figure 72). Correlation coefficients and curve shapes were used to assess the suitability of this technique with regard to processing by flow cytometers employing digital signal processing. At r^2 = 0.9544, 0.9703, 0.9545 and 0.941 correlation coefficients determined for MEFL, MEPE, MEPCY MEAP did not reach the anticipated >0.99 value specified by

Spherotech. All curves also display non-conformance to anticipated straight-line profiles, exhibiting a tendency toward convex (MEFL, MEPE, MEPCY) and convex (MEAP) curvature.

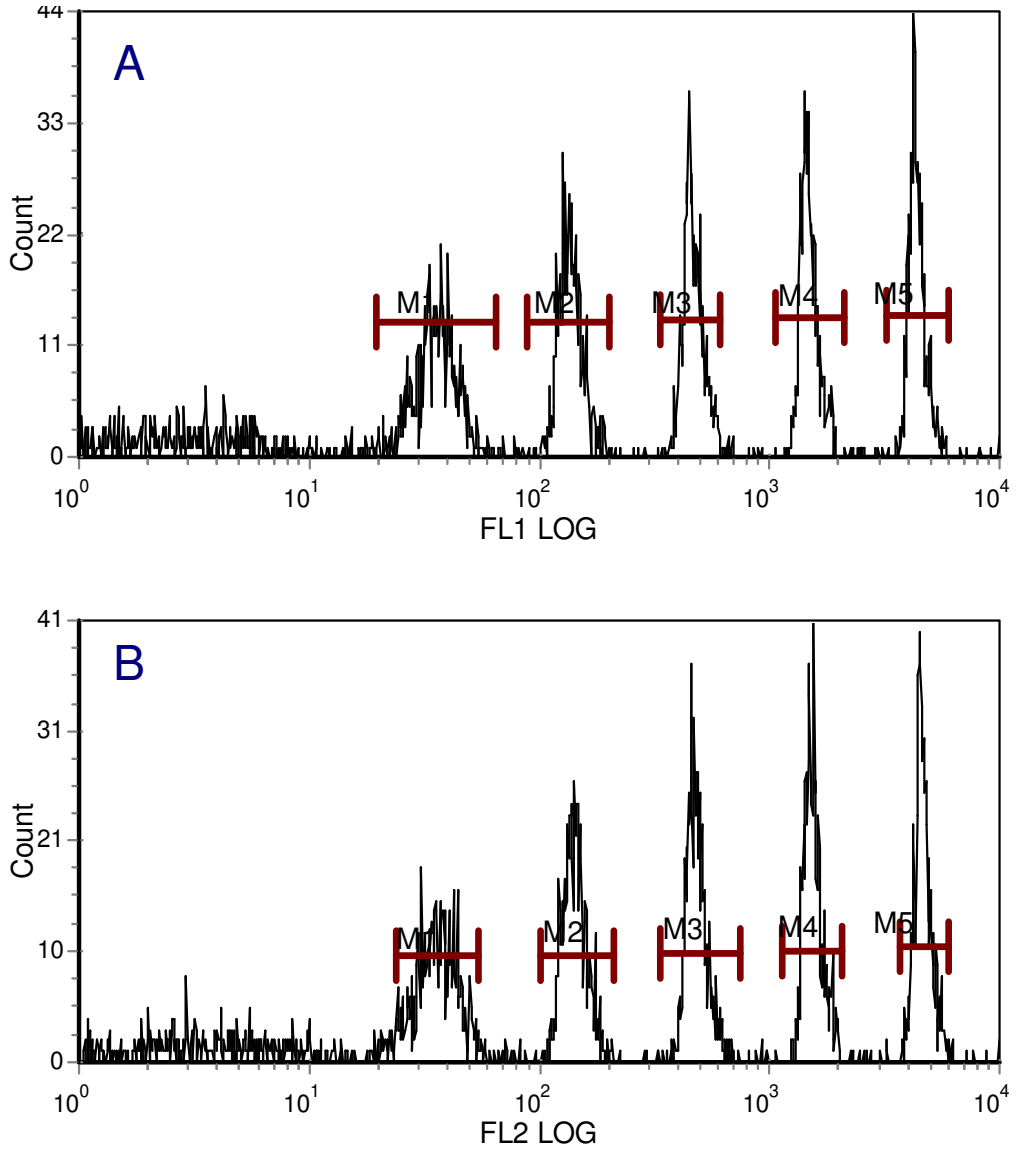


Figure 69: Histogram plots of Spherotech UltraRainbow calibration particle emission profiles in FITC (525nm) and PE (570nm) channel. Five populations are visible spanning four log decades; peaks M1-M5 are used for molecules of equivalent fluorescein (MEFL) and molecules of equivalent phycoerythrin (MEPE) determination.

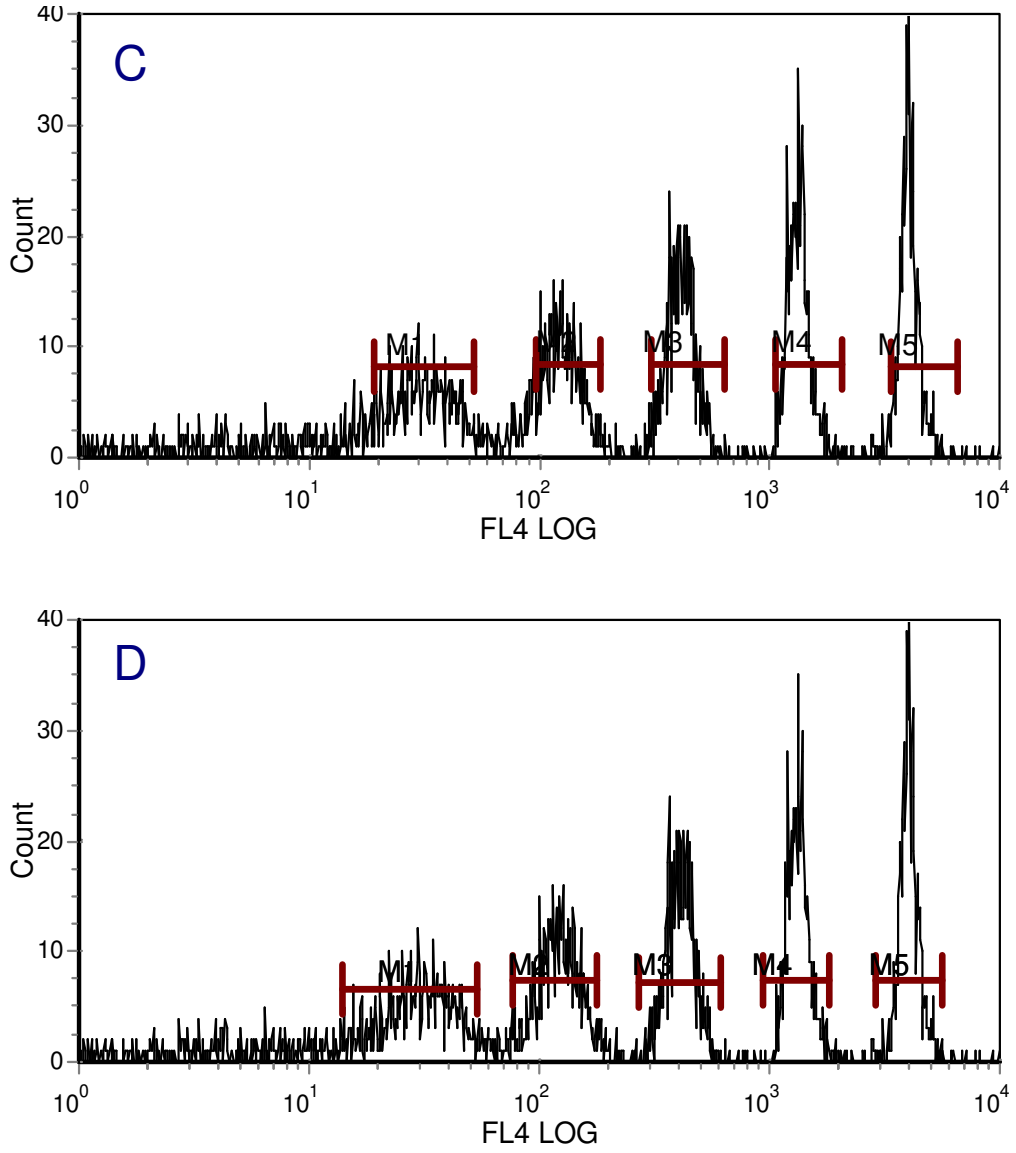
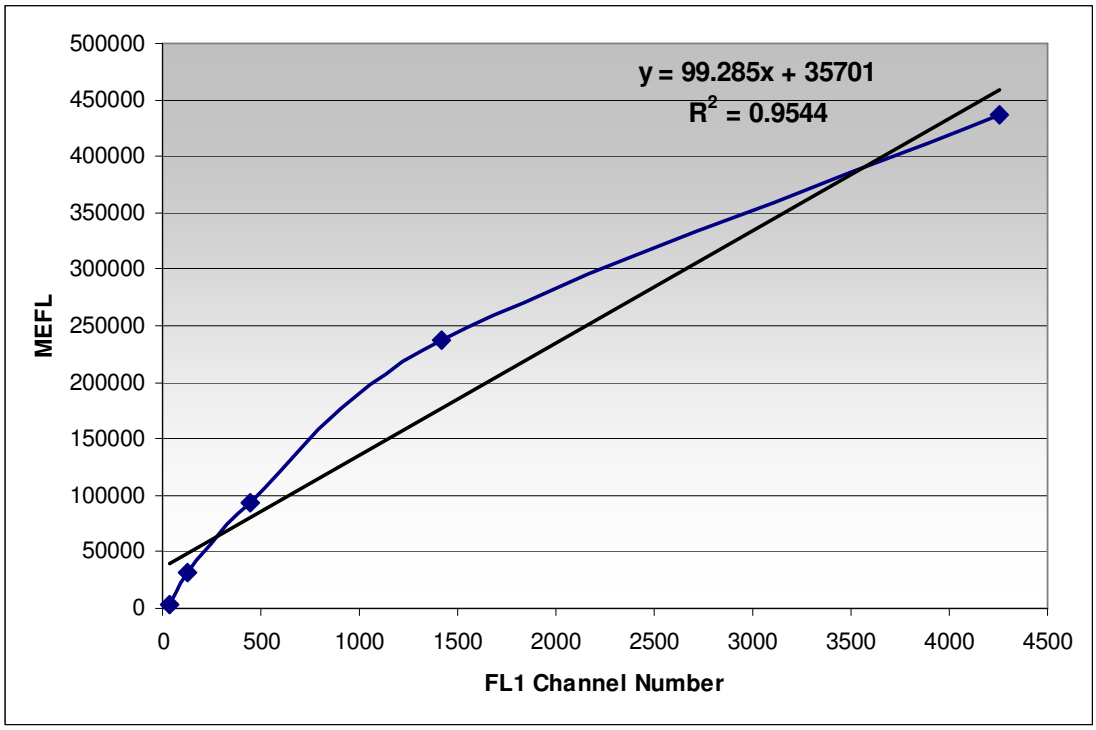


Figure 70: Histogram plots of Spherotech UltraRainbow calibration particle emission profiles in PE, PE-Cy5 / APC channels (histograms B, C and D respectively). Five populations are visible spanning four log decades. Peaks M1-M5 are used for determination of molecules of equivalent phycoerythrin (570nm), phycoerythrin-Cy5 and allophycocyanin (675nm).



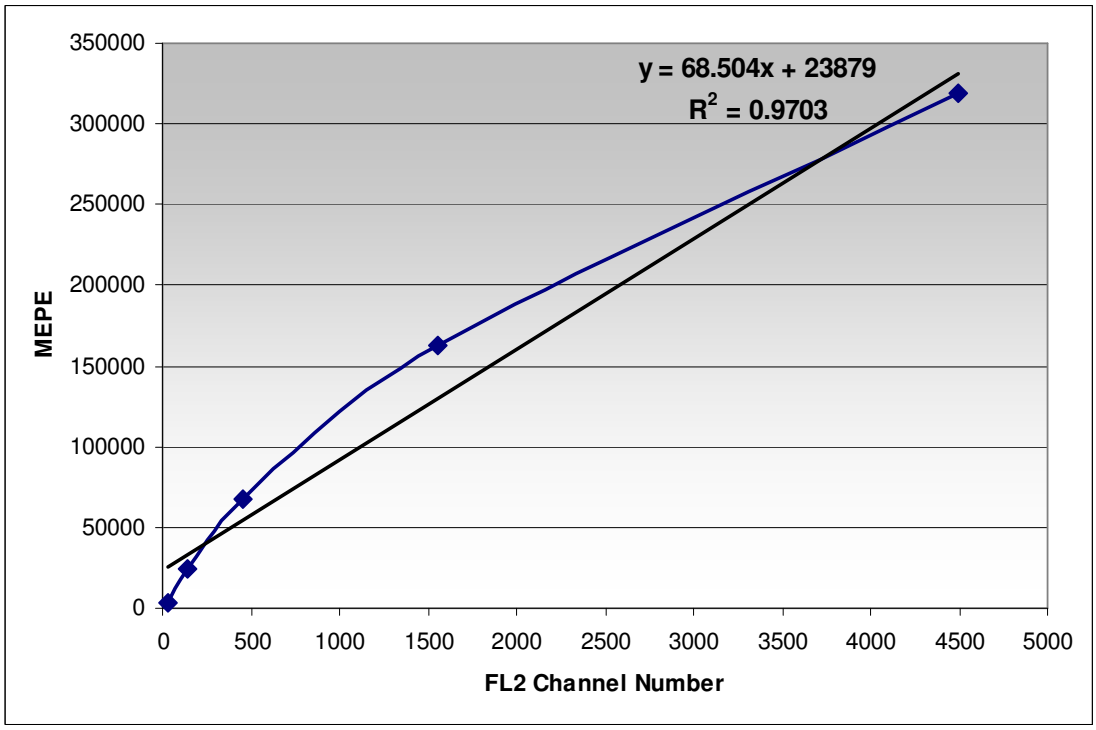


Figure 71: Calibration curves determined for MEFL and MEPE channels (525 and 575nm respectively) using EPICS XL derived channel numbers and Spherotrech supplied molecules of equivalent fluorescein and phycoerythrin. Correlation coefficiencts determined for both MEFL and MEPE curves (r2 = 0.9544 and 0.9703 respectively) are suboptimal and to not reach the anticipated >0.99 value specified by Spherotech. Both curves also display non-conformance to anticipated straight-line profiles, exhibiting a tendency toward convex curvature.

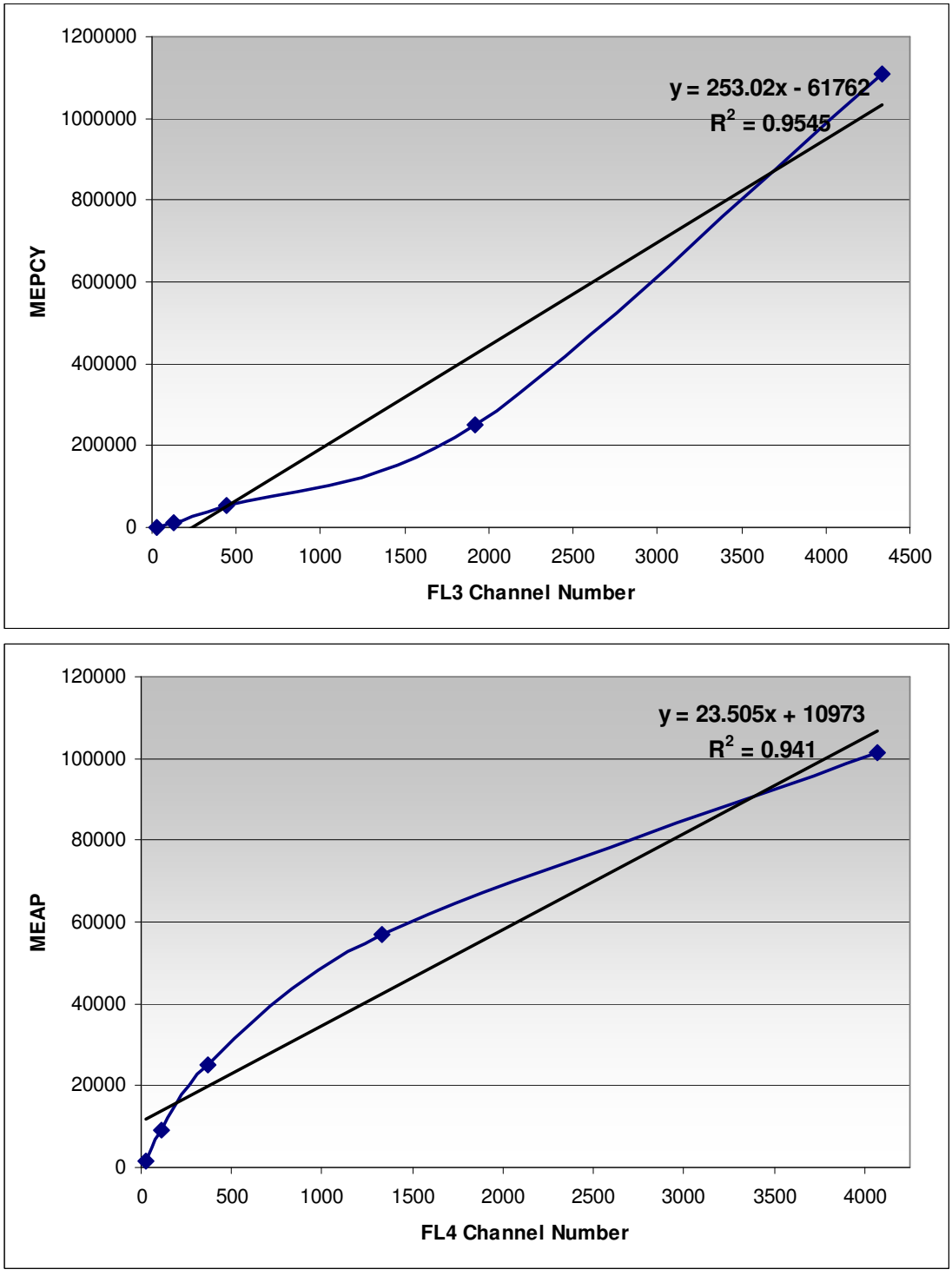


Figure 72: Calibration curves determined for MEPCY and MEAP channel (675nm) using EPICS XL derived channel numbers and Spherotrech supplied molecules of equivalent phycoerythrin-Cy5 and allophycocyanin. Correlation coefficiencts determined for both MEPCY and MEAP curves (r2 = 0.9545 and 0.941 respectively) are suboptimal and to not reach the anticipated >0.99 value specified by Spherotech. Both curves also display non-conformance to anticipated straight-line profiles, exhibiting a tendency toward concave and convex curvature for MEPCY and MEAP respectively.

5.3.3 Coupling Efficiency

Oligonucleotide probe immobilisation was facilitated using heterobifunctional EDC (1-Ethyl-3-[3-dimethylaminopropyl]) cross-linking between carboxylated microspheres and aminated oligonucleotide probes. A synthetic PE-Cy5 labelled aminated-Poly(dA) reporter probe was used to facilitate coupling optimisation. A number of factors including probe concentration and incubation duration were assessed to facilitate maximum coupling of probe sequences (Figure 74 and Figure 75 respectively). Microsphere analysis was performed using EPICS XL flow cytometric analysis and geometric mean determination used to assess coupling efficiency. A final probe concentration of 1pmol/µl (molar excess of 624:1 reporter probe molecules to COOH) in association with two hours incubation with rotation resulted in maximum coupling and nMFI of 121.25 (Figure 73).

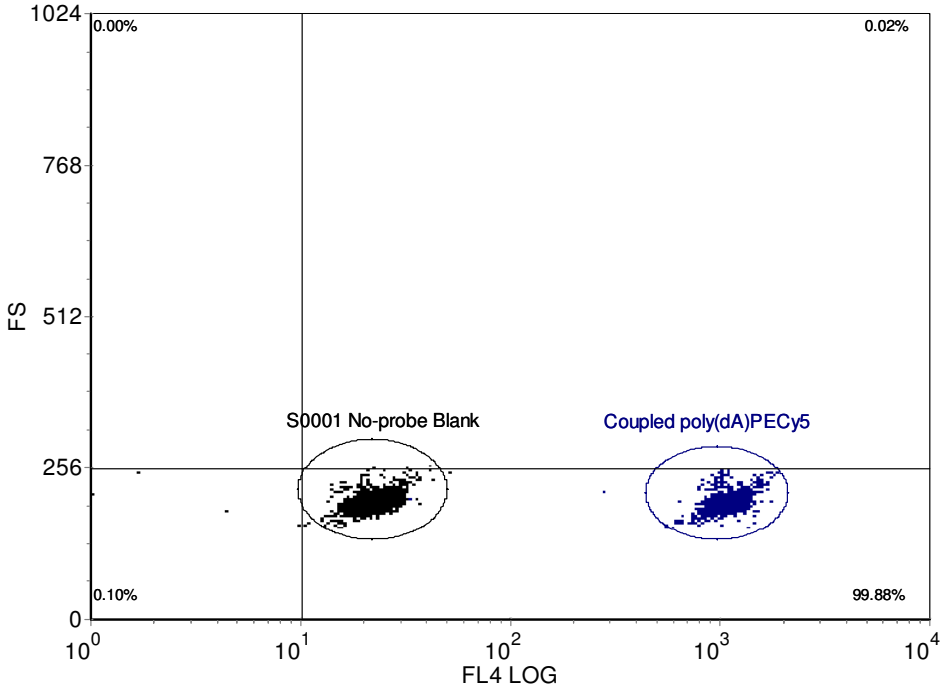


Figure 73: Coupling efficiency determination. Microsphere S0001 and PE-Cy5 labelled amino-Poly(dA) reporter probe were used to perform probe coupling optimisation. The FCS Express derived dot plot displays an overlay of two samples; a no-probe S0001 population and a S0001 reporter probe coupled population. A normalised geometric MFI of 103.91 was determined for the optimised coupled population

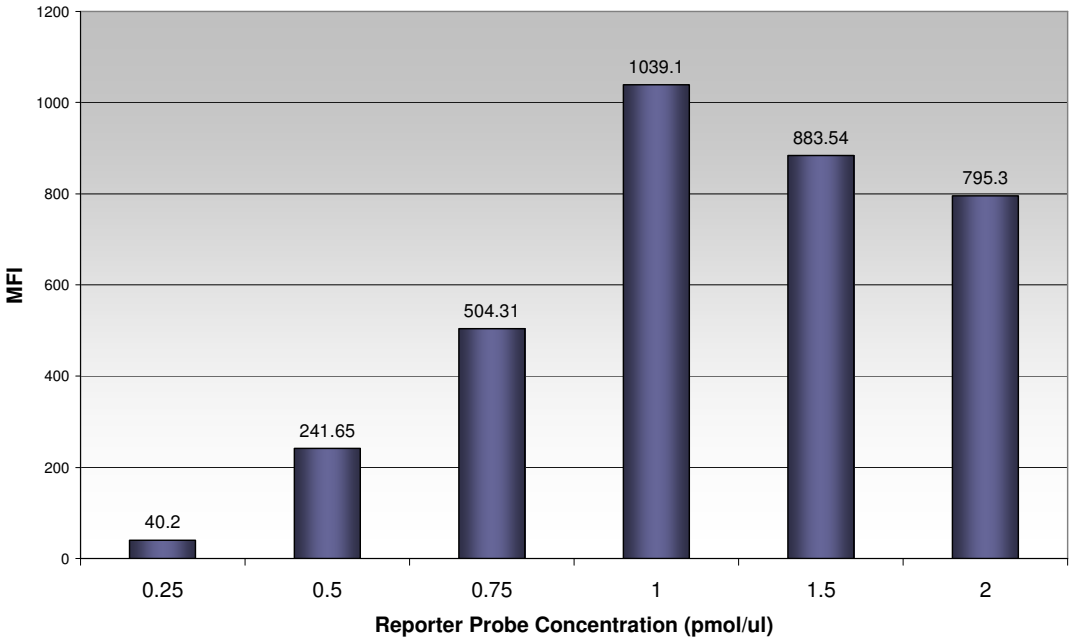


Figure 74: Effect of reporter probe poly(dA)PECY5 on coupling efficiency. A range of reporter probe concentrations ranging from 0.25-2.00pmol/µl were used to identify the optimal probe concentration required for coupling 2.9×10^4 microspheres. A final probe concentration of 1pmol/µl resulted in maximum coupling. This is a large molar excess of 624:1 reporter probe molecules to COOH- sites.

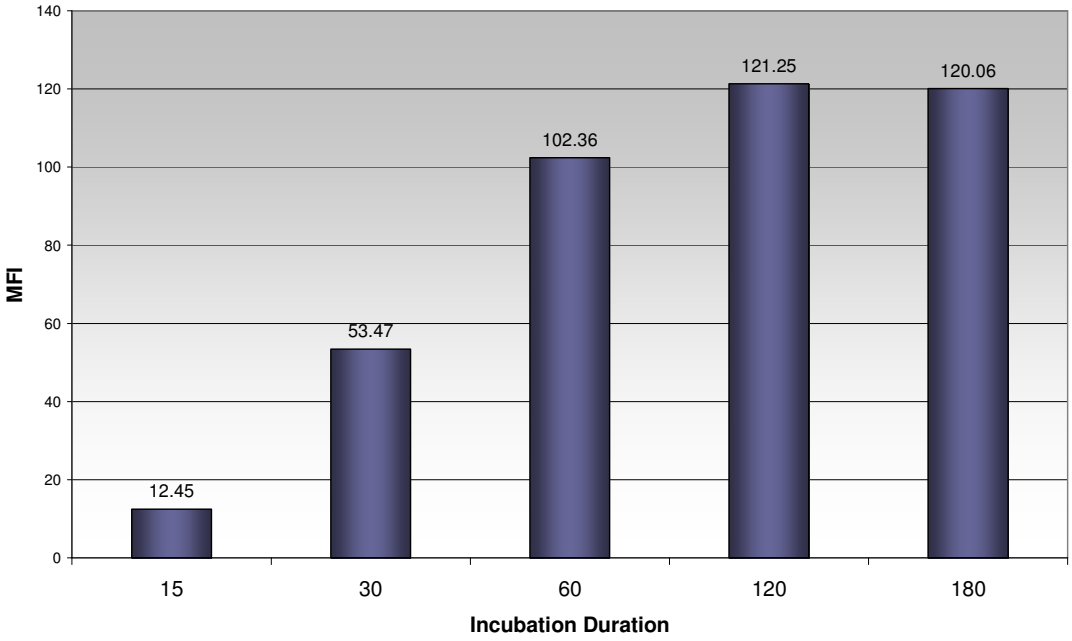


Figure 75: Effect of Incubation duration on coupling efficiency. MES buffered EDC carbodiimide cross-linking was used to perform coupling between S0001 microspheres and PE-Cy5 labelled reporter probe. Coupling efficiency was assessed following incubation durations of between 15 minutes and three hours. Two hours incubation resulted in MFI of 121.25 and maximum coupling efficiency.

5.3.4 Hybridisation Efficiency and Allele Designation

Hybridisation reactions were performed in accordance with the protocol described and a number of optimisations performed to allow maximum discrimination for genotyping (Figure 76).

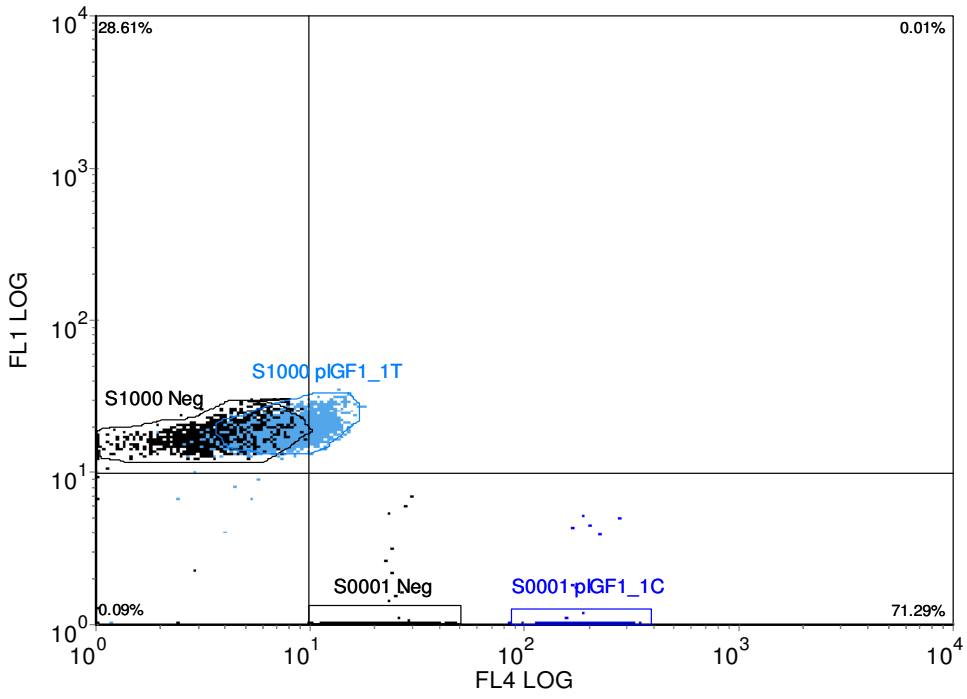


Figure 76: FCS Express dot plot for allele discrimination determination. All S1000 microspheres are coupled with pIGF1_1T probes and S0001 microspheres coupled with pIGF1_1C probs. Populations S1000 Neg and S0001 Neg highlighted in black were hybridised with non-complementary IGF1_5 singleplex PCR targets (i.e. target negative) while populations S1000 pIGF1_1T and S0001 pIGF1_1C highlighted in light blue and navy respectively were hybridised with IGF1_1 singleplex PCR target. Fully complementary allele positive S0001_pIGF1_1C / IGF1_1 species display a high normalised MFI of 157.64, this is 42.72 times higher than that displayed by single allele mismatch S1000_pIGF1_1T/IGF1_1 species which display a normalised MFI of 3.69 relative to their non-complementary target negative counterparts.

Hybridisation temperatures of 40-70°C spanning the predicted probe annealing temperature of 60°C were analysed and their effect on allele discrimination evaluated (Figure 77). Temperatures of 60°C were found to facilitate maximum allele discrimination capacity with normalised MFI of 157.64 and 3.69 determined for positive pIGF1_1C and negative pIGF1_1T allele probe-targets respectively. Lower hybridisation temperatures provided increased total MFI for both probes however divergence between

probe MFIs was reduced and allele discrimination was not facilitated at lower 40°C temperatures.

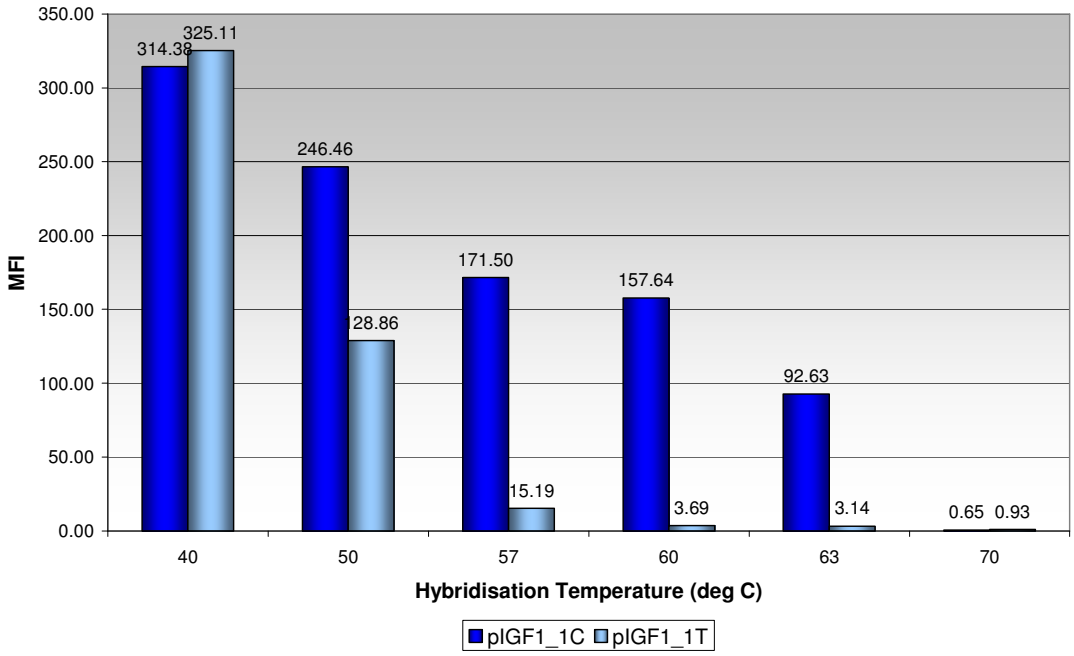


Figure 77: Hybridisation temperature optimisation. The effect of hybridisation temperature on allele discrimination was assessed across a 30°C range of 40-70°C. Temperatures of 60°C were found to facilitate maximum allele discrimination capacity with MFI of 157.64 and 3.69 determined for positive pIGF1_1C and negative pIGF1_1T allele probe-targets respectively. Lower hybridisation temperatures provided increased total MFI for both probes however divergence between probe MFIs was reduced and allele discrimination was not facilitated at lower 40°C temperatures. The highest 70°C hybridisation temperature exceeds the predicted T_M and as expected did not facilitate hybridisation (normalised MFI of 0.65 and 0.93 for pIGF1_1C and T respectively).

The effect of hybridisation duration was also assessed. One-hour hybridisation resulted in maximum normalised MFI of 161.97. Incubation times exceeding this did not act to improve hybridisation further (Figure 78).

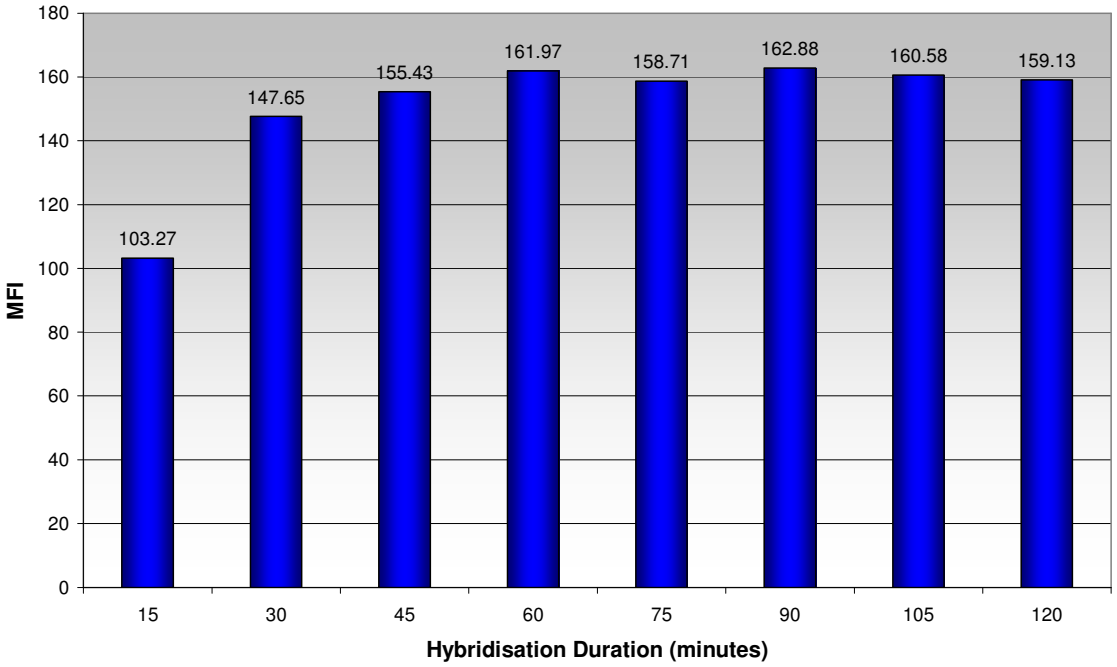


Figure 78: Effect of incubation duration on hybridisation efficiency. Using fully complementary probe and singleplex target (pIGF1_1C and IGF1_1) hybridisation was performed at 60°C with rotation over a series of incubation times ranging from 15 minutes to two hours. Sample MFIs were plotted against hybridisation duration to identify the optimal hybridisation time under the conditions specified. One-hour hybridisation resulted in maximum normalised MFI of 161.97, times exceeding this did not act to improve hybridisation.

The effect of target concentration using both singleplex and multiplex (14-plex) targets were assessed. Target concentrations can affect allele discrimination and it is therefore essential to determine a range which may be used effectively [153]. It was determined that allele discrimination could be performed effectively using either singleplex IGF1_1 or mltiplex (14-plex) target across a range of 5-50fMols. Use of multiplex targets reduced total MFIs somewhat however this effect was relatively small and allele discrimination was not impinged. Singleplex target allele discrimination spanning 5-50fmols of target for both IGF1_1C and IGF1_1T are displayed in Figure 79. Singleplex and multiplex (14-plex) MFIs determined across a range of concentrations from 5-50fmols for probe IGF1_1C are displayed in Figure 80, while those for probe IGF1_1T are displayed in Figure 81.

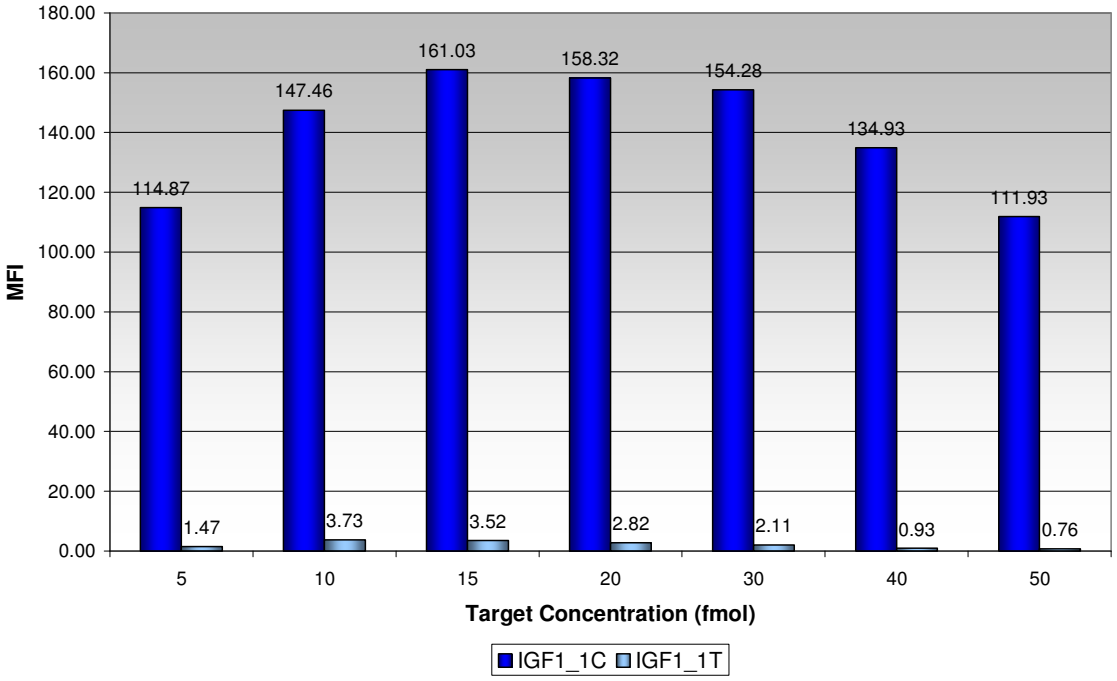


Figure 79: Effect of target concentration on allele discrimination. Hybridisation was performed using a range of singleplex IGF1_1 target concentrations from 5-50 fetomols. Normalised MFIs ranging from 114.87-161.03 and 0.76-3.73 were determined for positive and negative allele probes (pIGF1_1C and pIGF1_1T respectively). Target saturation was reached at 15 fmols for complementary pIGF1_1C samples, while a lower saturation point at 10fmols and poorer relative performance at concentrations exceeding this point were evident for allele negative pIGF1_1T samples.

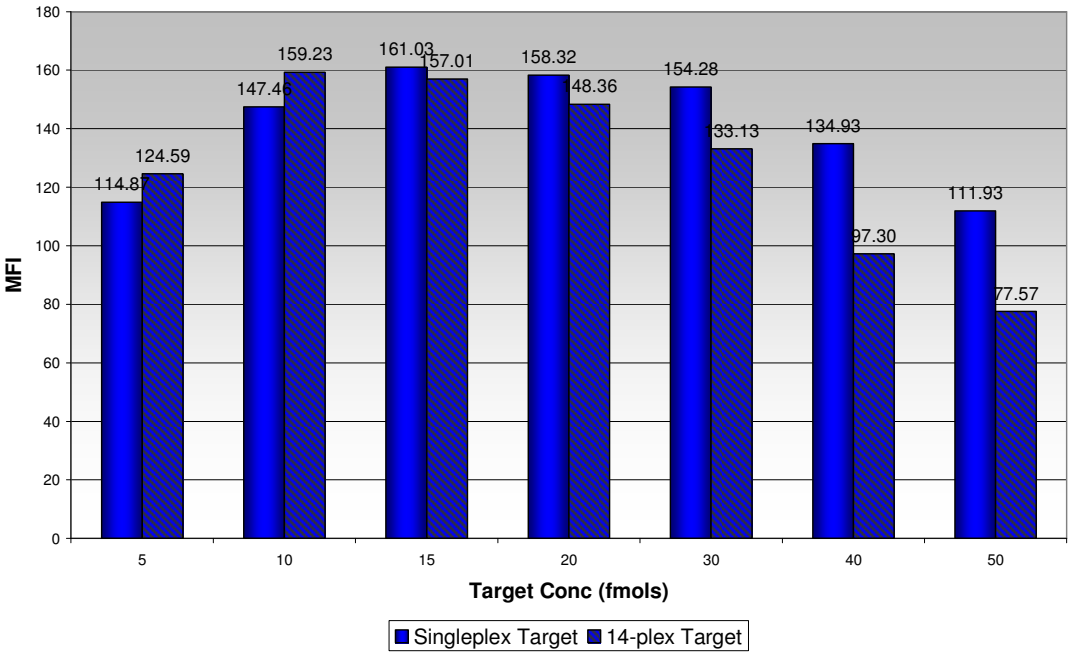


Figure 80: Effect of PCR target complexity on pIGF1_1C Hybridisation Efficiency. Singleplex IGF1_1 and 14-plex PCR product (containing IGF1_1 amplicon) were used to assess the performance of pIGF1_1C on hybridisation efficiency over a range of target concentrations (5-50fmols). Use of 14-plex targets resulted in a lower saturation point of 10fmols relative to that determined for singleplex targets (15fmols). Use of 14-plex target also resulted in marginally reduced overall performance.

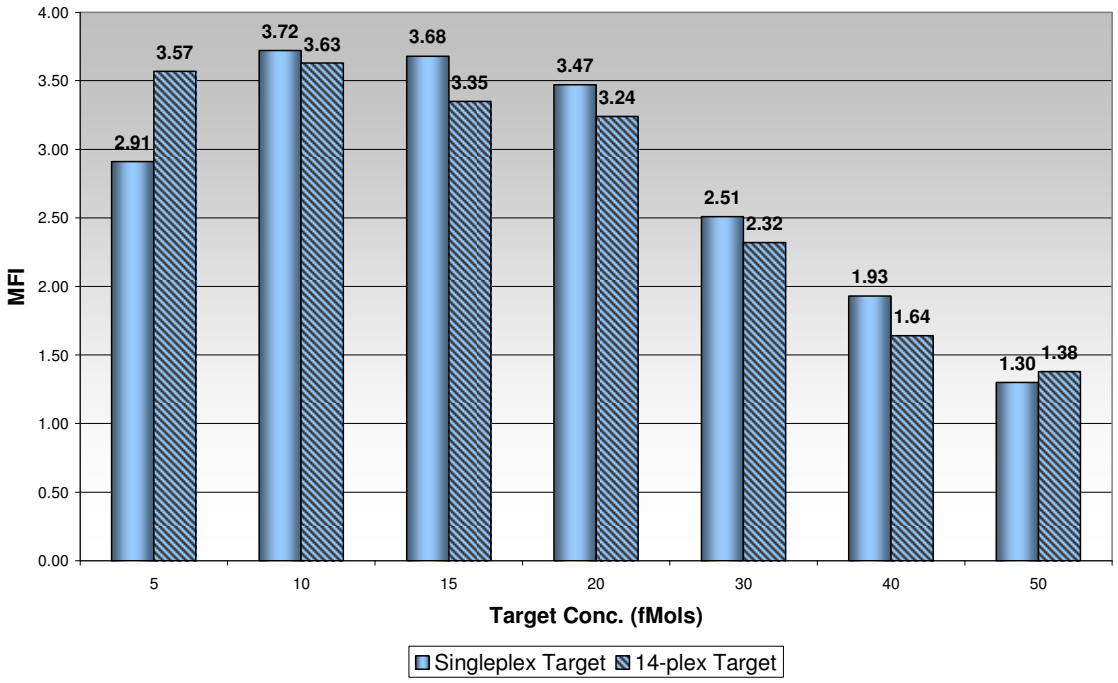


Figure 81: Effect of PCR target complexity on pIGF1_1T Hybridisation Efficiency. Singleplex IGF1_1 and 14-plex PCR product (containing IGF1_1 amplicon) were used to evaluate the performance of pIGF1_1T on hybridisation efficiency over a range of targets (5-50fmols). A target saturation point of 10fmols was determined for both singleplex and multiplex targets however multiplex targets displayed increased relative hybridisation signal at lower 5fmol concentrations.

5.3.5 Dot Blot Validation

Allele discrimination was also performed using dot blot hybridisation to validate results determined using suspension array approaches. Linear probe sequences (designed for suspension array genotyping) were hybridised with PCR product under a range of high-stringency conditions and hybridisation signal strength used to determine alleleotype of target sequences. Three probes were hybridised in parallel with singleplex IGF1_1 PCR amplified target. Strongest hybridisation signals were seen for IGF1_1C in all cases, a lower relative hybridisation signal was seen for all IGF1_1T samples bar the highest stringency wash D where no signal was evident. No hybridisation signal was observed using IGF1_5G probe negative control.

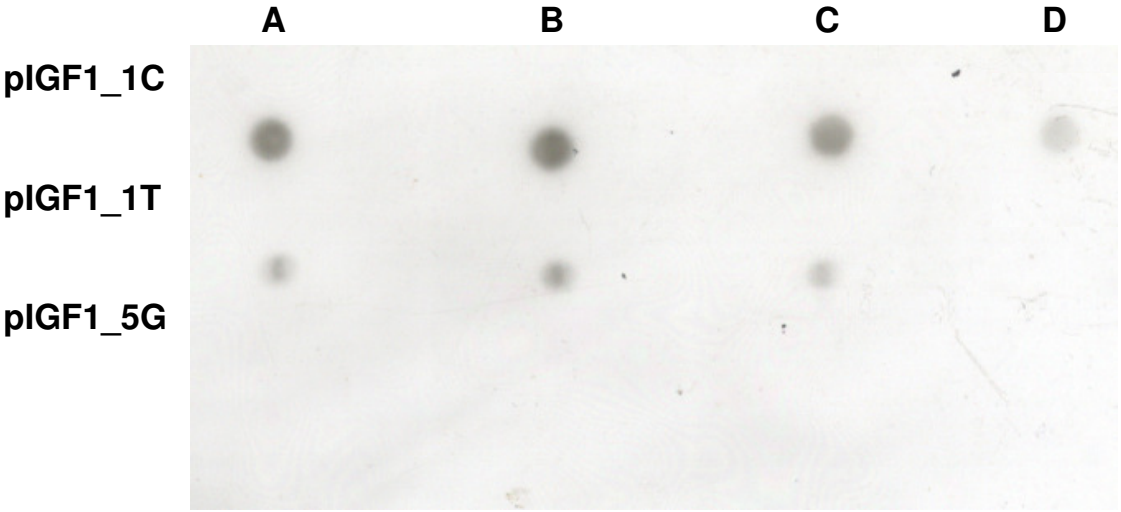


Figure 82: Dot blotting was used to validate allele designation as determined by suspension array experimentation. IGF1_1 singleplex PCR product was chemically denatured and spotted onto nitrocellulose blotting membrane. IGF1_1C, IGF1_1T and IGF1_5G probes were used to assess the target sequence using alternate primary washes increasing in stringency from 0.2-0.5X SSC from A to D respectively. Strongest hybridisation signals were seen for IGF1_1C in all cases, a lower relative hybridisation signal was seen for all IGF1_1T samples bar the highest stringency wash D where no signal was evident. No hybridisation signal was observed using IGF1_5G probe even with the lowest stringency wash A.

5.4 Discussion

Suspension array technology offers the potential for high throughput SNP genotyping of PCR products with high accessability, flexability, sensitivity, speed and efficiency [13]. Like any oligonucleotide assay however all probe sequences require extensive optimisation and validation prior to clinical sample analysis. A wide range of alternate approaches with regard to suspension array coupling and hybridisation have been demonstrated. The following section describes development of a high efficiency suspension array method for use in conjunction with high-dimensionality multiplex PCR targets and probe sequences as previously described (chapters 3 and 4). Microsphere selection, coupling, hybridisation and validation of a model probe pair pIGF1_1T and pIGF1_1C in conjunction with singleplex and multiplex PCR targets are described demonstrating the efficacy of the suspension array genotyping method developed.

5.4.1 Microsphere Characterisation and Selection

In order to exploit the inherent advantages afforded by quantum dots (narrow emission peaks with minimal red tailing and high capacity for multiplexing); suspension array solid supports composed of quantum dot encoded microspheres were selected. A number of companies including Bangs Laboratories and Duke scientific supply low dimensionality combinatorial sets of single colour, multi-intensity carboxyl-functionalised vinyl polymer microspheres of this nature, however Crystalplex offered the highest dimensionality combinatorial set in suitably discrete quantities in the form of the PlxBead nanocrystal encoded microspheres [230,275].

Particle counts were performed using haemocytometer and flow cytometry, demonstrating a mean PlxBead concentration of $1.46x10^4/\mu l$ determined for all beads in the set (although the concentration of microsphere QDEM 1100 was substantially below that of other microspheres at $1.7 \times 10^1/\mu l$). Moderate to significant microsphere loss was noted upon Initial coupling and subsequent hybridisation experiments however and as such, potential contributing factors including buffer composition and concentration, salt concentration, and pH were investigated in an attempt to determine the cause of microsphere atrophication.

The effect of H_2O, TE, PBS and TMAC (2.5M) storage on microsphere populations were tested at intervals over the course of two days. Water storage had no discernable effect while significant microsphere loss due to TE storage within the assay time span was manageable. Stark effects were seen following PBS and TMAC incubation however with significant loss observed following two hours and thirty minutes PBS and TMAC incubation respectively. PBS has a low salt concentration (similar to that of TE buffer) as such osmotic shock seemed unlikely and the reasons for this effect were investigated. Consultation with the manufacturers revealed phosphate to be utilised in PlxBead manufacture and subsequent storage in phosphate-containing buffer was found to cause leeching of constituent quantum dots and microsphere disruption (personal correspondence Crystalplex). The effect of TMAC was also of concern, TMAC is a soluble organic salt which binds selectively to AT rich regions reducing melting temperature disparity between AT and GC rich regions, as such facilitation of this salt was essential for multiplex probe analysis of the nature proposed [276]. Incubation of PLxBeads with 2.5M TMAC caused significant loss following of microspheres following one hour incubation. Salt-induced osmotic shock was proposed as a potential mechanism

by which atrophication may be induced and therefore a range of TMAC concentrations (0.01-4M) were investigated. Microsphere population attrition of 10% was observed in solutions of <1M, however attrition increased significantly thereafter with ~80% loss determined following incubation with 4M TMAC solutions. TMAC solutions were also adjusted for pH evaluation with micropsheres appearing to be most stable in higher pH solutions (up to pH 10).

Alternate hybridisation protocols such as those described by Das et al., which facilitate hybridisation using lower salt buffers, may potentially be used in conjunction with methacrylate PlexBead microspheres. However, use of lower salt conditions significantly reduces hybridisation kinetics incurring increased hybridisation time requirements [277,278]. Additionally one of the advantages afforded by suspension array technology is that microsphere coupling may be performed in bulk and aliquots subsequently used to perform numerous genotyping experiments [13]. It was felt that microsphere instability was too significant to proceed using microspheres as described and therefore amelioration techniques and alternate probe supports were investigated.

PlxBeads are composed of mesoporous methacrylate (PMMA), a clear plastic resin favoured for its robust characteristics. However PMMA can be unstable and manufacturing must be carefully controlled if robust structures are to be produced [279]. Core-shell deposition has been the subject of intense investigation with silica, alumina and titania investigated in a bid to increase core shell stability [280]. Silica based core-shell hybrids are the most widely applied due to their low cost, chemical inertia and transparency which renders then suitable to a wide range of chemical and biological applications [280,281]. A number of treatment options including microsphere surface

modification in association with coupling agent employment as well as electrostatically driven layer-by-layer self assembly for sequential adsorption of silica have been employed for vinyl polymer encapsulation. However the requirement for surface treatment and repeated centrifugation / wash / redispersion treatments make these processes time consuming [280-284]. Cao *et al*., successfully employed the former approach to coat carboxylated polystyrene QD encoded beads using Tetraethylorthosilicate (TEOS) and demonstrate functionality of reactive groups through DNA probe hybridisation, however the low pH (pH1.5) employed in this instance was found to be incompatible with methacrylate PlxBead stability and as such an NTC/nanosilica approach which facilitates electrostatic deposition of silica on PMMA as developed by Chen *et al*., was identified as a potential solution to the PLxBead stability problem [280,281]. These stabilisation issues were discussed with Crystalplex and development of silica-encapsulated Plxbeads is currently underway. In the interim however alternate supports were sought.

Although optically encoded using fluorophores (rather than quantum dots) Spherotech particles were selected due to their reported stability with regard to the wide salt and pH ranges which may be accommodated (personal correspondence Spherotech) [285,286]. Microsphere stability in terms of resistance to buffer composition and concentration, salt concentration, and pH were again tested and microspheres found to be highly-stable inducing insubstantial microsphere loss under all conditions tested. Two Spherotech microspheres species S1000 and S0001 with maximum emission at 530nm and 635nm as described were thus used for all subsequent suspension array genotyping optimisation experiments.

5.4.2 Flow Cytometry Data Analysis

Suspension array coupling and hybridisation procedures performed during the course of this work use fluorescent reporter molecules to facilitate assay evaluation. A number of approaches, including absolute and relative quantitative analysis, exist to allow meaningful determinations regarding genotyping outcomes to be made.

Quantitative flow cytometry facilitates quantitative measurement of particle staining by flow cytometry, providing an absolute value for the light intensity measured. Using microsphere sets containing multiple populations with known fluorophore loadings, mean channel numbers may be determined for specific instruments / assays allowing standard curve construction and absolute fluorophore determination for unknown samples [228]. A number of commercial calibration kits including Quantum^{TM} PE-Cy5 MESF, QuantiBRITE-PE and UltraRainbow Calibration Particles have been specially devised to allow accurate repeatable determinations to be made in this way [228,230,287]. The Spherotech UltraRainbow Calibration kit was selected due to its multichannel emission spectra and use of water insoluble fluorescent dyes which afford increased stability relative to surface labelled microparticle attachment methods used in other calibration kits. Their high size uniformity also means that these species may be used for optical system alignment [288,289].

The applicability of this approach to our EPICS XL instrument was assessed using URFP-38-2K UltraRainbow Calibration Particles which should allow molecules of equivalent FITC, PE, PE-CY5 and APC to be determined in channels FL1, FL2 and FL4 using our optical set-up. FC parameters were selected to allow detection of calibration particle across four log decades and calibration particles analysed. Derived channel numbers were plotted against known molecules of equivalent fluorochrome (MEF) and

calibration curves plotted for each channel. Spherotech calibration particles have very low coefficients of variation and should display correlation coefficients of >99% (personal correspondence Spherotech), however deviation from anticipated profiles were noted for MEFL, MEPE, PEPCY and MEAP standard curves which displayed correlation coefficients of 0.9544, 0.9703, 0.9545 and 0.941 respectively. Non-conformance of data points to the anticipated straight-line profile for was noted for all curves with MEFL/MEPE/MEAP and MEPCY standard curves displaying convex and concave curvature respectively. The reasons for these deviations are unknown however the EPICS XL employs digital signal processing which uses linear amplifiers in association with log look-up tables to convert linear signals to logarithmic output. All other commercially available flow cytometers use analog based methods which employ log amplifiers for signal amplification. These amplifiers are notoriously difficult to align however and as such true logarithmetic output is difficult to obtain [265]. UltraRainbow Calibration Particles were developed and optimised using 12 cytometers including Dako Cyan, various LSRs, FAC StarPlus, Canto I, FC500, and MoFlo which implement use of log amplifiers for signal processing. As such amplification induced deviation may have contributed to the deviant profiles observed in this instance (personal correspondence, Spherotech) [228].

An alternate data interpretation approach involves use of mean fluorescent intensity determination (MFI). Using this measure of central tendency, fluorescent profiles for alternate populations or treatment types can be measured to allow relative fluorescence determination to be made. Although quantitative flow cytometry should facilitate empirical characterisation of fluorophore loading and therefore more objective characterisation, error is likely to be induced by use of non-conforming standard curves.

MFIs have been used extensively for suspension array analysis of this nature with Xu *et al.,* Armstrong *et al.,* and Yeoma *et al.,*using this method for single nucleotide discrimination [10,153,290], as such MFI determination was selected to facilitate efficiency determination and allele discrimination in this instance [291,292].

5.4.3 Microsphere Coupling

Oligonucleotide probes may be immobilised on the microsphere surface using non-covalent physical adsorption, affinity binding methods, or covalent conjugation. Physical adsorption of DNA onto glass particles occurs due to electrostatic interaction, however the process is not very stable (subject to pH induced deviation) or efficient and it also prohibits the use of the newer optically encoded high-dimensionality multiplex sets which are generally composed of hydrophobic polymers. High-affinity tags such as poly-his, biotin and glutathione-S-transferase may be coupled with their respective ligands to produce stable, high-specificity reactions, however in terms of both of these parameters, covalent coupling using carboxylated microspheres and aminated oligonucleotides offers the greatest gains and is the most widely used oligonucleotide attachment approach currently employed [13,237,293].

Heterobifunctional cross-linker EDC was used to perform carbodiimide coupling between carboxyl groups and primary amines. During this reaction an *O*-acylisourea intermediates are formed however these groups are unstable and susceptible to hydrolysis and as such this reaction can be difficult to perform effectively [293,294]. A number of amended protocols have been developed in an attempt to improve coupling efficiency including a high-pH (7.4) imidazole buffer and low-pH MES (pH 4.5) facilitated approach which can be performed using a single step or multiple step [10,272,290,295,296]. These approaches have all been applied efficiently however lower-

pH MES facilitates faster reaction rate and was therefore selected for attachment of target-specific probe sequences in this instance [295]. Recent studies by Xu *et al*., have incorporated a poly(dT) linker sequence for coupling evaluation, however this requires hybridisation of fluorescent poly(dA) target sequences and as such an extra source of variability with regard to hybridisation efficiency and specificity is included [10]. We employed an aminated-poly (dA) probe directly labelled with PE_Cy5 fluorophore to facilitate optimisation of the coupling reaction.

Stoichiometric calculations using manufacturer derived data determined the surface of each 5µM microspheres to be coated with approximately 3.32256×10^6 carboxyl sites, with an inter-carboxyl distance of ~2,284$Å^2$ and carboxyl group to inter-carboxyl surface area of 1:29. This should allow binding of the reporter probe in sufficient quantities to facilitate detection [296].

A range of reporter probe concentrations ranging from 0.25-2.00pmol/µl were used to identify the optimal probe concentration required for coupling 2.9 × 104 microspheres. A final probe concentration of 1pmol/µl resulted in maximum coupling with normalised MFI of 103.91 following two hours incubation. This is a large molar excess of 624:1 reporter probe molecules to COOH- sites, mirroring that determined by Newkirk *et al*., which cites an optimum ratio of 600:1 probe to potential binding sites [272].

5.4.4 Hybridisation and Allele Designation

The optimised coupling protocol was used to perform coupling between target-specific probe sequences IGF1_1T and IGF1_1C and microspheres S0001 and S1000. Hybridisation between microsphere-probe moieties and PCR products were then performed. The effect of applied hybridisation temperature is stark especially with regard to allele discrimination where conditions must be stringent enough to facilitate identification of heterozygous genotypes. Lower hybridisation temperatures have been implemented to perform allelic designation of this nature however stringency washes must be applied to remove un-specifically bound target and total MFI signals are reduced. As such the effect of hybridisation temperature on allele discrimination was assessed across a 30°C range of 40-70°C. Optimal hybridisation temperature designation was made by identifying the temperature at which complementary duplexes produced an increased MFI such that discrimination between homozygous positive and negative alleles was highest while retaining a population that was clearly distinguishable from target-negative controls. Un-reacted species must be clearly distinguishable from duplexed species if allelic determination is to be robust to coupling/hybridisation aberrations.

With normalised MFIs of 161.03 and 3.69 (C and T probe alleles respectively) and relative fluorescence increase 43.63 times that for homozygous positive relative to homozygous negative alleles, optimal hybridisation temperature was designated to be 60°C for probe pair IGF1_1. Lower hybridisation temperatures provided increased total MFI with normalised MFIs of 314.38 and 322.52 for pIGF1_1T and C respectively determined at 40°C, in the order of MFI determined for reporter probe coupled species,

however divergence between alternate allele probe MFIs was reduced and allele discrimination was not facilitated. The highest 70°C hybridisation temperature exceeds the predicted T_M and as expected did not facilitate hybridisation (mean MFI = 0.79).

Blotting has been performed by a number of groups to validate suspension array alelelotyping [10]. Dot blotting was also performed using singleplex IGF1_1 target and probes in this instance. Significantly stronger hybridisation signals were determined for membranes treated using probe pIGF1_1C relative to those derived using pIGF1_1T, using the highest stringency wash (containing 0.5xSSC) no signal was determined for probe IGF1_1T thereby validating previously determined SAT genotyping results.

Total nMFI determined using this method at optimal hybridization temperature was lower than that determined by a number of other groups including Dunbar *et al.,* who determined net MFIs at ~800 [274] . Individual probe kinetics and melting temperatures are likely to contribute to variations between MFIs for alternate probes however spacer length has also been found to contribute heavily to total MFI magnitude. Spacers are employed to reduce stearic hindrance and increase hybridisation efficiency Using a number of linear probe and PCR target sequences Shchepinov *et al*., determined that employment of 40 atom modified-nucleotide spacers increased hybridisation signals by up to 150 times [167]. A number of additional spacer types including standard nucleotide and carbon atom spacers have been investigated, however repulsive forces of negatively charges nucleotides have been shown to reduce hybridisation while both carbon spacers and modified nucleotides are expensive [297]. A balance between spacer length and cost must therefore be struck. 12-carbon spacers are most often employed for suspension based genotyping however a 6-Carbon spacers was employed during the course of this

work [167,277]. Use of this shorter spacer type may have contributed to reduced total MFI however it may also acted to increase discrimination between alleles (to make any determination of this nature however genotyping using a more extensive probe set is required).

The hybridisation kinetics and thermodynamic affinities of nucleic acid duplexes can be driven in a concentration dependant manner. Increasing target concentrations are known to improve MFI signals, however an excess of target can result in poor allele discrimination and overall signal reduction due to renaturation competition between complementary PCR stands [298]. The effect of singleplex target concentration was investigated using IGF1 PCR amplicon targets ranging from 5-50fmols. MFIs ranging from 161.03-111.93 and 3.79-0.76 were determined for positive and negative allele species (C and T alleles respectively) across this range with saturation point of 15 fmols determined for fully complementary sequences. A lower saturation point of 10fmols was determined for species containing homozygous mismatches.

It may be postulated that increasing PCR mixture complexity could affect hybridisation efficiency and allele discrimination. The effect of PCR target complexity on pIGF1_1T hybridisation was therefore assessed over a range of target concentrations (5-50fmols) using singleplex IGF1_1 and 14-plex PCR product (as detailed in chapter 4). Use of higher complexity 14-plex targets resulted in lower overall performance. Lower saturation points were also incurred relative to those determined singleplex targets, with maximum target capacities of 10fmols determined for complementary and single-base mismatch species respectively.

It is important to note that SAT probe saturation points are not merely a product of the hybridisation protocol employed. Target concentration ranges are also influenced by the innate profile and associated hybridisation rate constants of individual target sequences used, as such use of relatively equimolar multiplex PCR targets during multiplex suspension array genotyping should provide a degree of leeway with regard to accommodation of this feature and reduce or negate the need for further multiplex PCR optimisation during SAT analysis [153].

The work in this chapter aimed to meet the final project objective "*To demonstrate the applicability of multiplex PCR to suspension array facilitated allele discrimination using a model probe pair and validate allele designation via dot blotting*". These aims were met and allele discrimination demonstrated using a range of both singleplex and multiplex (14-plex) PCR targets. Allele designation was made and validated using dot blotting.

Chapter 6: Overall discussion, conclusions and further work

6.1 Overall Discussion

The work presented in this thesis was produced in order to satisfy several key objectives which were defined to allow the aims to be met. The aim of the thesis was to "*use haplotype based tagSNP selection and a systematic in silico-based analysis approach to design a multiplex compatible PCR primer and SAT probe set facilitating maximum variation capture with minimum tests across candidate genes IGF1, IGFBP1 and IGFBP3. This will be applied though development of a number of robust, high-efficiency, high-specificity multiplex PCR constructs for amplification of multiple targets to demonstrate the applicability of these target types to suspension array genotyping for NIDDM*".

The first objective was "*to construct haplotypes for a given population and perform tagSNP selection which captures maximum variation across candidate genes IGF1, IGFBP1 and IGFBP3*".

Linkage disequilibrium based tag-SNP selection was used to select target SNPs across the candidate loci. Haplotypes were constructed by Haploview, using population data of predominantly Caucasian origin, and TagSNPs selected using Tagger. Due to the block-based structure of the genome and instigation of a MAF threshold of ≤5%, a significantly reduced SNP population may be selected for analysis with little reduction in power making larger associations more cost-effective and amenable.

Using the method as described a total of 292 SNPs spanning candidate genes IGF1_1, IGFBP1 and IGFBP3 were identified. Of these 83 displayed sufficient minor allele frequencies of ≥5% and were sectioned into a total of 32 bins. IGF1_1 contained 39 SNPs

in 13 bins capturing alleles with a mean r^2 of 0.973, IGFBP1_1 contained 19 SNPs partitioned in five bins displaying a mean r^2 of 0.957, while IGFBP3 contained 25 SNPs were partitioned into 14 bins displaying a mean *r2* of 0.968. All tag and capture SNPs displayed an r^2 exceeding 0.8, indicative of strong linkage disequilibrium between all alleles. Linkage disequilibrium was seen to be strongest across IGF1_1 and weakest across IGFBP3, reflected by the bin to SNP ratio determined to be 1:3 and 1:1.5 for IGF1_1 and IGFBP3 respectively. Through prioritisation of SNPs displaying high linkage disequilibrium and moderate minor allele frequencies, LD-based tagSNP selection facilitated knowledge-based SNP prioritisation, reducing the number of SNPs required for comprehensive IGF1_1, IGFBP1 and IGFBP3 gene coverage from 292 to 32 with minimal information loss.

All tag and capture SNPs were also investigated by PupaSuite in order to identify those located in transcription factor binding sites, triplex-forming oligonucleotide target sequences, intron/exon border consensus sequences, exonic splicing enhancers and exonic regions, with putative deleterious effects which may contribute to the NIDDM disease profile. Although selection of SNPs was not made on the basis of putative functional impact, it was felt that downstream experimental incompatibility may require incomplete bin representation which could be informed by results of such analyses.

Using PupaSuite, IGF1 candidate tagSNPs; rs6214 / rs6219 and IGFBP3 capture SNP rs2854746 received deleterious impact predictions. IGF1 and IGFBP3 SNPs were found to be located in SRp40 responsive and SR Sf2 responsive ESEs representing a potential propensity toward mRNA processing misfunction [112]. Functional studies regarding the nature of the putative deleterious SNPs identified during this study have not been

performed to date however in association with supportive evidence (i.e. strong association), these loci may also represent good targets on which to perform functional analysis (knock-down, selective mutation studies etc.).

Following successful knowledge-based SNP reduction and bin prioritisation, the second objective "*To perform multiplex primer and probe design, utilising in silico and manual analysis for evaluation and selection of a high specificity primer/probe sets in a manner compatible with downstream multiplex PCR and SAT analysis*" could be addressed.

While use of increased dimensionality formats can reduce costs; design and optimisation of these complex structures is considerably more problematic than for singleplex formats. The probability of successful multiplex analysis performance may be ameliorated by extensive and comprehensive *in silico* analysis however, and to this end Primer3, UCSC PCR, Primer Map and AutoDimer were used to design multiplex PCR and SAT compatible primers and probes for maximum bin coverage. While adherence to singleplex design recommendations was important; use of multiplex PCR formats is known to incur diminished tolerance ranges and induce the potential for cross-reactivity. As such, a number of multiplex-specific primer design features were implemented, allowable thresholds restricted, and multiplex-specific tests performed to increase the probability of successful amplification.

Deviations from optimal T_A can affect efficiency of amplification, which can be exasperated within multiplex formats resulting in large yield variances between alternate amplicons. Narrow inter- and intra-primer T_A ranges of $\leq 2°C$ were therefore stipulated for this study. Using the Panjkovich consensus T_M method which displays minimal T_M

error probabilities, all primers in the final IGF multiplex set adhered to this range with maximum inter-primer T_A of just 1.68°C and intra-primer pair T_A of 1.32°C [141]. A second feature of increased primer length was also implemented to aid specificity. Primer length is generally observed in the range of 18-24 bases for amplification of human genomic targets. Sanchez *et al.,* and Henegariu *et al.,* suggested that use of longer primers (up to 35 bases) and their associated elevated T_As may increase specificity for multiplex amplification [162,299]. These principles were applied for primer design in this instance resulting in production of primers with an average length of 27 bases ranging from 22-30 nucleotides.

AutoDimer analysis was also performed to assess cross-reactivity of all primer and probe species which may reduce assay performance. AutoDimer scores in excess of 13 were determined for self-tagging rs33979592, rs2854744 and rs2453840 IGFBP3 SNPs primer pairs (bins two, five and eight respectively). This exceeds the maximum recommended score threshold of seven to eight and as such all three tagSNPs were excluded from further analysis. Cross-homology constraints did not affect PupaSuite functional variants and therefore prioritization was not made on this basis.

A number of relaxed parameters were allowed for a number of primer pairs where superior tagSNP replacement options were unavailable. Primer pairs exceeding these optimal constraints include pairs IGF1_9, IGF1_10 and BP3_14, located in repeat regions known to increase susceptibility to non-specific amplification. IGF1_12 and IGF1_8, which at 500bp and 78bp respectively, exceed the optimal 100-300bp size range. IGF1_9 which failed the *in silico* PCR test, and may therefore contain homologous targets despite the longer primer lengths used. In addition to IGF1_10's less than optimal target type,

these primers were also constructed without stabilising G/C clamps which may further exasperate reaction specificity.

Probe sequences also displayed high T_M conformity with intra- and inter-probe T_Ms ranging a maximum of 1.32°C and 3.49°C, As such, probe T_M-induced genotyping irregularities should be minimised. Genotyping efficiency may be influenced by a number of probes including IGF1_2 and IGF1_11 which, due to sequence constraints were designed with non-central target SNPs, a feature known to reduce SNP-induced duplex instability. Due to position along the complementary sequence of probes pBP1_3 and pBP1_4, SNP rs9658195 may interfere with genotyping. However this is predicted to have a low MAF frequency and further genotyping of this position in the target population is required. Consequently, primer and probe sequences representing 29 of 32 bins across targets genes were designed during this phase.

While *in silico* evaluation can act to improve the probability of high-level primer performance, it is no guarantee of experimental success, particularly for those primers exceeding optimal specifications. As such the objective described for chapter 4 "*To perform extensive PCR optimisation for the construction of a number of robust, well characterised, high dimensionality multiplex PCR sets"* aimed to address this and produce targets for downstream SAT applications.

Singleplex end-point PCR was performed to assess primer pair specificity. All primer pairs displayed target specific amplification bar repeat region pairs IGF1_9, IGF1_10 and BP3_14 which displayed multiple band formation. A number of polymerases, adjuvants and reaction protocols were used to reduce non-specific amplification; sufficient (target-

specific) amelioration was achieved for IGF1_9 however, only when using high-fidelity Ampli*Taq* gold polymerase protocols.

According to Panjkovich T_A prediction; all primers in the final IGF multiplex set should have an optimal T_As of between 64.47 and 65.81°C (mean T_A = 65.24°C), however prediction agreement using this method is strongest for oligonucleotide sequences <20–22 nucleotides in length. A mean primer length of 27 nucleotides (ranging from 22-30bases) primers in our multiplex set exceeded high-specificity prediction limits. As such RT-PCR optimal T_A assessment was performed across five annealing temperatures ranging from 59.9-68.8°C. Despite the relatively long primer lengths used, the Panjkovich method displayed good predictive performance with 71% or primer pairs functioning optimally at the predicted T_A (~65°C) [141]. A slight tendency towards underestimation of T_A was noted; with 25% of primer pairs (including IGF1_5, IGF1_12, BP1_3/4, BP1_5, BP3_3, BP3_7) functioning optimally at T_A = 67.5°C. Just one primer pair (BP3_10/11) displayed optimal performance at T_A = 62.2°C. Relative yield analysis determined that all primers (excluding low-specificity IGF1_9, IGF1_10 and BP3_14) functioned well at 65.0°C, therefore this annealing was selected for amplification of multiplex constructs.

PCR amplification efficiency reflects primer pair performance and is indicative of stable high-specificity yield production, while dynamic range reflects the ability of primer pairs to function optimally across a wide range of target concentrations and is important for multiplex amplification where tolerances are known to diminish. As such RT-PCR was performed for all primer pairs across a dynamic range spanning four orders of magnitude.

Twenty four high-specificity primer pairs displayed good amplification efficiency[13] across this dynamic range with mean efficiency, slope and correlation coefficients of 101.90%, -3.27 and 0.988 respectively. These are within the ideal efficiency (90-100%), slope (-3.6 to -3.1) and correlation coefficient (>0.95) limits generally recommended. Primer pair IGF1_8 displayed poor efficiency and a suboptimal slope, however at 72bp, IGF1_8 is produces the shortest amplicon, therefore, its requirement in terms of optimal extension times is below that of other pairs. Use of inappropriate annealing time may have contributed to the poor performance observed in this instance.

The results of these singleplex amplification procedures were used to inform end-point multiplex set construction and amplification [224,225]. Primer pairs IGF1_10 and BP3_14 primer pairs, shown during non-specific amplification under all conditions tested, were removed from the set. Repeat region IGF1_9, shown to be susceptible to non-specific artifact formation, was purposely assembled into the lowest dimensionality 2-plex construct to minimise complications imposed by more complex mixtures and Ampli*Taq* gold-derived protocol (previously shown to facilitate target-specific amplification of IGF1_9) used in place of the Qiagen multiplex mix. In addition to this 2-plex structure, a high-specificity, high efficiency 14-plex, 6-plex and 5-plex were produced for amplification of a total of 27 targets across this 99kb region. To the best of our knowledge; the larger 14-plex IGF construct developed during the course of this work is among the highest dimensionality multiplexes assembled for targeted gene-specific disease association with 14 primer pairs spanning ~99kb (~1 test / 7kb). Thi sis exceeded however by Makowski *et al's.,* 31-test cystic fibrosis multiplex spanning 188kb which facilitates one test every ~6kb [300].

[13] Excluding low-specificity IGF1_9, IGF1_10 and BP3_14 and high-specificity IGF1_8

Multiplex amplification for disease association appears to be quite limited with many associations utilising singleplex or low complexity multiplex amplification (<5-plex) [251,301-304]. The largest multiplex of this nature identified during this review was an 8-plex developed by Mirel *et al.,* for IDDM based association testing [305]. To date, disease association has been primarily based on assessment of putative functional polymorphisms which cannot be substituted. Exploiting the haplotypic structure of genes for SNP prioritisation can facilitate inter-bin tagSNP substitution where alternate tags are unavailable, thereby allowing greater gene coverage to be retained. The target regions used during this study were relatively small, spanning 84.7kb, 5.3kb and 9.0kb across candidate genes IGF1, IGFBP1 and IGFBP3 respectively. The high linkage disequilibrium displayed by IGF1 resulted in accumulation of numerous interchangeable tagSNPs rendering block representation relatively straight forward for standard non-repeat region targets. Adversely IGFBP3, which displayed the lowest low levels of linkage disequilibrium, allowed little room for manoeuvre in terms of tagSNP substitution. As a result lower multiplex coverage was more challenging with non-repeat region tagSNPs rs33979592, rs2854744 and rs2453840 (IGFBP3 bins two, five and eight respectively) unrepresented by any primer pair.

The PCR products developed during this work were used to attempt to satisfy the final objective "*To demonstrate the applicability of multiplex PCR to suspension array facilitated allele discrimination using a model probe pair and validate allele designation via dot blotting*"

Originally we proposed the use of quantum dot encoded microspheres which, due to their narrow emission spectra and multi-wavelength excitation, should facilitate large multiplex set construction to meet the demands of the 58 probe set. Methacrylate microspheres were sourced from Crystalplex (Crystalplex Corp. Pittsburgh, PA), however these unprotected methacrylate microspheres were found to be unstable and susceptible to breakage especially in the presence of high-salt solutions such as those used during our hybridisation assays. Potential amelioration approaches were researched and the results of our findings discussed with the manufacturer. As a result a more robust, silica-encapsulated species is currently under development.

In the interim however alternate, more stable microspheres were sourced; we selected two fluorophore-encoded polystyrene microspheres produced by Spherotech (Spherotech, Inc., Lake Forest, IL) Although encoded using fluorophores, these highly-stable microspheres facilitated a suitable level of multiplex required for demonstration of our model 2-plex SAT genotyping assay. Optimisation of carbodiimide EDC-coupling was achieved using a synthetic PE-Cy5 labelled aminated-Poly(dA) reporter probe and the optimised protocol applied to facilitate high-efficiency model probe-pair IGF1_1C and IGF1_1T attachment. Hybridisation was subsequently performed and conditions optimised to allow SNP discrimination. OE21 cell line was determined to display homozygous CC allele at the target locus (rs972936). The applicability of both singleplex IGF1_1 and 14-plex target (containing IGF1_1) were assessed and good efficiency and discrimination determined for both target types. The total MFI was somewhat reduced using the multiplex target structure and saturation point was reached at an earlier juncture, however overall, the homozygous discrimination thresholds were generally in-

line with those described by other groups [153,260,306]. Dot blot experiments validated genotyping results determined using SAT analysis.

The applicability of these multiplex PCR targets to biallelic suspension array-based genotyping has been demonstrated, paving the way for development of a larger multiplex suspension array using the probe set designed during the course of this work. Although currently untested, the success of *in silico* primer design, de-risks multiplex probe construction, particularly in terms of its likely propensity for cross-homology. The multiplex nature of the PCR amplicon set described should allow association of candidate genes IGF1, IGFBP1 and IGFBP3 to be performed at a significantly reduced cost (in European Caucasian populations). As such, a larger proportion of association analysis budgetary provision may be allocated towards analysis of larger sample sizes, thereby increasing the power of association outcomes. The increased accessibility of this set may also encourage the scope of association to be expanded allowing knowledge-driven and exploratory associations to be performed across the polygenic disease spectrum.

While the applicability of our IGF multiplex targets was demonstrated by use of high throughput bi-allelic suspension array based genotyping, further development of the full probe set may allow high-dimensionality analysis of multiplex targets to be performed thereby reducing the test structure for both amplification and genotyping tests (with associated cost reduction which would be incurred). Additionally the multiplex PCR amplicons developed during the course of this work may also be analysed using a number of alternate genotyping approaches making this multiplex structure accessible to the wider research community.

6.2 Overall Conclusion

The aim of this work was to "use haplotype based tagSNP selection and a systematic in silico-based analysis approach to design a multiplex compatible PCR primer and SAT probe set facilitating maximum variation capture with minimum tests across candidate genes IGF1, IGFBP1 and IGFBP3. This will be applied though development of a number of robust, high-efficiency, high-specificity multiplex PCR constructs for amplification of multiple targets to demonstrate the applicability of these target types to suspension array genotyping for NIDDM".

This has been addressed through several objectives which have previously been discussed. From the work presented here, several conclusions may be drawn:

- Haplotype based tagSNP selection for European Caucasian populations has been performed and a systematic *in silico* PCR primer and SAT probe design and analysis approach implemented to thoroughly assess all target tagSNPs allowing a well considered, multiplex-compatible structure capturing maximum variation with minimum tests across candidate genes IGF1, IGFBP1 and IGFBP3 to be designed.

- A set of 25 primer pairs, capturing 27 target bins have been profiled thoroughly using singleplex PCR amplification. From these four robust multiplex-compatible PCR sets (14-plex, 6-plex, 5-plex and 2-plex) which display high-efficiency and target-specific amplification have been developed.

- The applicability of these multiplex PCR constructs to suspension array genotyping has also been demonstrated using a single probe pair, thus paving the

way for development of a large multiplex suspension array genotyping assay using the probes designed during the course of this work.

As such a more accessible lower test, higher-efficiency analysis structure has been developed which facilitates comprehensive association analyses of IGF1, IGFBP1 and IGFBP3 to be performed.

6.3 Further work

To test all probe pairs designed in chapter 3 to determine their allele discrimination capacity and construct these into large multiplex sets which facilitate analysis of multiplex PCR targets.

To test the full multiplex capacity of all high-specificity primer pairs to increase the dimensionality of our multiplex set using fluorescent amplified fragment length polymorphism (fAFLP). FAFLP uses fluorophores rather than intercalating dye in combination with capillary electrophoresis for size dependant separation of multiplex amplicons. As such fluorochromes with alternate emission wavelengths can be utilised for identification of identically sized amplicons irresolvable by single colour detection. Our primer set is readily amenable to this type of analysis as the biotinylated attachment implemented for downstream SAT can also be utilised for streptavidin facilitated fluorophore attachment without any further adjustment [307].

We would like to apply the techniques (multiplex PCR - suspension array method) developed during this thesis to perform an association analysis to determine whether any association between IGF1, IGFBP1 and IGFBP3 SNPs captured by these targets and NIDDM is evident in European Caucasian populations.

Reference List

1. Palmer L.J., Cordon L.R. Genetic Epidemiology 4 - Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* 2005; 366:1223-1234.

2. Hsiao A., Kuo M.D. High-throughput biology in the postgenomic era. *Journal of Vascular and Interventional Radiology* 2006; 17:1077-1085.

3. Jonathan L.Haines, Margaret Pericak-Vance Genetic Analysis of Complex Diseases . 2007.

4. GPUSDEOS. Human Genome Project Information - SNP Fact Sheet. Genomes Programs of the U.S Department of Energy Office of Science (GPUSDEOS) . 19-9-2008. 3-11-0008.

Ref Type: Electronic Citation

5. Bhatti P., Church D.M., Rutter J.L., Struewing J.P., Sigurdson A.J. Candidate single nucleotide polymorphism selection using publicly available tools: A guide for epidemiologists. *American Journal of Epidemiology* 2006; 164:794-804.

6. Carlson C.S., Eberle M.A., Rieder M.J., Yi Q., Kruglyak L., Nickerson D.A. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; 74:106-120.

7. de Bakker P.I.W., Yelensky R., Pe'er I., Gabriel S.B., Daly M.J., Altshuler D. Efficiency and power in genetic association studies. *Nature Genetics* 2005; 37:1217-1223.

8. Gu S., Pakstis A.J., Kidd K.K. HAPLOT: a graphical comparison of haplotype blocks, tagSNP sets and SNP variation for multiple populations. *Bioinformatics* 2005; 21:3938-3939.

9. Edwards M.C., Gibbs R.A. Multiplex Pcr - Advantages, Development, and Applications. *Pcr-Methods and Applications* 1994; 3:S65-S75.

10. Xu H., Sha M.Y., Wong E.Y. et al. Multiplexed SNP genotyping using the Qbead system: a quantum dot-encoded microsphere-based assay. *Nucleic Acids Res* 2003; 31:e43.

11. M.J.Fulwyler. Method for detecting and separating antigens and antibodies in blood or other samples. [1561042]. 1976. UK.

Ref Type: Patent

12. Horan P., Wheeless L. Quantitative Single Cell Analysis and Sorting. *Science* 1977; 198:149-157.

13. Nolan J.P., Sklar L.A. Suspension array technology: evolution of the flat-array paradigm. *Trends in Biotechnology* 2002; 20:9-12.

14. Ronald Aubert Diabetes in America. *DIANE Publishing,* 1995.

15. R A S Hemat Principles of orthomolecularism. *Urotext,* 2003.

16. Koster J.C., Permutt M.A., Nichols C.G. Diabetes and insulin secretion - The ATP-sensitive K+ channel (K-ATP) connection. *Diabetes* 2005; 54:3065-3072.

17. Butler A.E., Janson J., Bonner-Weir S., Ritzel R., Rizza R.A., Butler P.C. beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* 2003; 52:102-110.

18. Rhodes C.J. Type 2 diabetes - A matter of beta-cell life and death? *Science* 2005; 307:380-384.

19. Butler A.E., Jang J., Gurlo T., Carty M.D., Soeller W.C., Butler P.C. Diabetes due to a progressive defect in beta-cell mass in rats transgenic for human islet amyloid polypeptide (HIP rat) - A new model for type 2 diabetes. *Diabetes* 2004; 53:1509-1516.

20. Cerasi E. Insulin Deficiency and Insulin-Resistance in the Pathogenesis of Niddm - Is A Divorce Possible. *Diabetologia* 1995; 38:992-997.

21. Groop L.C., Widen E., Ferrannini E. Insulin-Resistance and Insulin Deficiency in the Pathogenesis of Type-2 (Non-Insulin-Dependent) Diabetes-Mellitus - Errors of Metabolism Or of Methods. *Diabetologia* 1993; 36:1326-1331.

22. Bloomgarden Z.T. Insulin resistance: Current concepts. *Clinical Therapeutics* 1998; 20:216-231.

23. Hales C.N. The Pathogenesis of NIDDM. *Diabetologia* 1994; 37:S162-S168.

24. Ostenson C.G. The pathophysiology of type 2 diabetes mellitus: an overview. *Acta Physiologica Scandinavica* 2001; 171:241-247.

25. Gerich J.E. The genetic basis of type 2 diabetes mellitus: Impaired insulin secretion versus impaired insulin sensitivity. *Endocrine Reviews* 1998; 19:491-503.

26. Reenders K., Denobel E., Vandenhoogen H., Rutten G., Vanweel C. Diabetes and Its Long-Term Complications in General-Practice - A Survey in A Well-Defined Population. *Family Practice* 1993; 10:169-172.

27. Mohan V., Vijayaprabha R., Rema M. Vascular complications in long-term south Indian NIDDM of over 25 years' duration. *Diabetes Research and Clinical Practice* 1996; 31:133-140.

28. Department of Health. About Diabetes. National Health Service . 18-5-2007. Ref Type: Electronic Citation

29. Diabetes UK. Diabetes numbers leap to 1.8 million people in the UK. 7-12-2004. Ref Type: Pamphlet

30. Alberto Pugliese, Donatella Miceli The Insulin Gene in Diabetes. *Diabetes Metab Res Rev* 2002; 18:13-25.

31. Qing-Yang HUANG, Meng-Rong CHENG, Sen-Lin JI Linkage and Association Studies of the Susceptibility Genes for Type 2 Diabetes. *Acta Genetica Sinica* 2006; 33:573-589 .

32. Rinderknecht E., Humbel R.E. Amino-Acid Sequence of Human Insulin-Like Growth Factor-I and Its Structural Homology with Proinsulin. *Journal of Biological Chemistry* 1978; 253:2769-2776.

33. Ullrich A., Gray A., Tam A.W. et al. Insulin-Like Growth Factor-I Receptor Primary Structure - Comparison with Insulin-Receptor Suggests Structural Determinants That Define Functional Specificity. *Embo Journal* 1986; 5:2503-2512.

34. Slaaby R., Schaffer L., Lautrup-Larsen I. et al. Hybrid receptors formed by insulin receptor (IR) and insulin-like growth factor I receptor (IGF-IR) have low insulin and high IGF-1 affinity irrespective of the IR splice variant. *Journal of Biological Chemistry* 2006; 281:25869-25874.

35. Moxham C.P., Duronio V., Jacobs S. Insulin-Like Growth Factor-I Receptor Beta-Subunit Heterogeneity - Evidence for Hybrid Tetramers Composed of Insulin-Like Growth Factor-I and Insulin-Receptor Heterodimers. *Journal of Biological Chemistry* 1989; 264:13238-13244.

36. Haluzik M., Yakar S., Gavrilova O., Setser J., Boisclair Y., LeRoith D. Insulin resistance in the liver-specific IGF-1 gene-deleted mouse is abrogated by deletion of the acid-labile subunit of the IGF-binding protein-3 complex - Relative roles of growth hormone and IGF-1 in insulin resistance. *Diabetes* 2003; 52:2483-2489.

37. Suryawan A., Nguyen H., Orellana R.A., Bush J.A., Liu C.W., Davis T.A. Insulin/insulin-like growth factor-I (IGF-I) hybrid receptor abundance decreases with development in neonatal pigs. *Faseb Journal* 2003; 17:A1158.

38. Clemmons D.R. The relative roles of growth hormone and lGF-1 in controlling insulin sensitivity. *Journal of Clinical Investigation* 2004; 113:25-27.

39. Yakar S., Liu J.L., Fernandez A.M. et al. Liver-specific igf-1 gene deletion leads to muscle insulin insensitivity. *Diabetes* 2001; 50:1110-1118.

40. Le Roith D., Bondy C., Yakar S., Liu J.L., Butler A. The somatomedin hypothesis: 2001. *Endocrine Reviews* 2001; 22:53-74.

41. Jones J.I., Clemmons D.R. Insulin-Like Growth-Factors and Their Binding-Proteins - Biological Actions. *Endocrine Reviews* 1995; 16:3-34.

42. Heald A., Stephens R., Gibson J.M. The insulin-like growth factor system and diabetes - an overview. *Diabetic Medicine* 2006; 23:19-24.

43. Bang P., Stangenberg M., Westgren M., Rosenfeld R.G. Decreased Ternary Complex-Formation and Predominance of A 29 Kda Igfbp-3 Fragment in Human Fetal Serum. *Growth Regulation* 1994; 4:68-76.

44. Moses A.C., Young S.C.J., Morrow L.A., OBrien M., Clemmons D.R. Recombinant human insulin-like growth factor I increases insulin sensitivity and improves glycemic control in type II diabetes. *Diabetes* 1996; 45:91-100.

45. Clemmons D.R., Moses A.C., Mckay M.J., Sommer A., Rosen D.M., Ruckle J. The combination of insulin-like growth factor I and insulin-like growth factor-binding protein-3 reduces insulin requirements in insulin-dependent type 1 diabetes: Evidence for in vivo biological activity. *Journal of Clinical Endocrinology and Metabolism* 2000; 85:1518-1524.

46. Clemmons D.R., Sleevi M., Busby W.H. Recombinant, nonglycosylated human insulin-like growth factor-binding protein-3 (IGFBP-3) is degraded preferentially after administration to type II diabetics, resulting in increased endogenous glycosylated IGFBP-3. *Journal of Clinical Endocrinology and Metabolism* 2005; 90:6561-6568.

47. Weinzimer S.A., Gibson T.B., Collett-Solberg P.F., Khare A., Liu B.R., Cohen P. Transferrin is an insulin-like growth factor-binding protein-3 binding protein. *Journal of Clinical Endocrinology and Metabolism* 2001; 86:1806-1813.

48. Lofqvist C., Chen J., Connor K.M. et al. IGFBP3 suppresses retinopathy through suppression of oxygen-induced vessel loss and promotion of vascular regrowth. *Proceedings of the National Academy of Sciences of the United States of America* 2007; 104:10589-10594.

49. Swanson M. Retinopathy screening in individuals with type 2 diabetes: who, how, how often, and at what cost—an epidemiologic review. *Journal of the American Optometric Association* 2005; 76:636-646.

50. Al Zahrani A., Sandhu M.S., Luben R.N. et al. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. *Human Molecular Genetics* 2006; 15:1-10.

51. Sandhu M.S., Heald A.H., Gibson J.M., Cruickshank J.K., Dunger D.B., Wareham N.J. Circulating concentrations of insulin-like growth factor-I and development of glucose intolerance: a prospective observational study. *Lancet* 2002; 359:1740-1745.

52. Ueki K., Okada T., Hu J. et al. Total insulin and IGF-I resistance in pancreatic beta cells causes overt diabetes. *Nature Genetics* 2006; 38:583-588.

53. Borecki I.B., Suarez B.K. Linkage and association: Basic concepts. *Genetic Dissection of Complex Traits* 2001; 42:45-66.

54. Andreas Ziegler, Inke R.König A statistical approach to genetic epidemiology. 2006.

55. Alpa V.Patel, Iona Cheng, Federico Canzian et al. IGF-1, IGFBP-1, and IGFBP-3 Polymorphisms Predict Circulating IGF Levels but Not Breast Cancer Risk: Findings from the Breast and Prostate Cancer Cohort Consortium (BPC3). *PLoS ONE* 2008; 3.

56. Dib C., Faure S., Fizames C. et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; 380:152-154.

57. Xing C., Gray-Mcguire C., Kelly J.A. et al. Genetic linkage of systemic lupus erythematosus to 13q32 in African American families with affected male members. *Human Genetics* 2005; 118:309-321.

58. Daw E.W., Heath S.C., Lu Y. Single-nucleotide polymorphism versus microsatellite markers in a combined linkage and segregation analysis of a quantitative trait. *BMC Genetics* 2005; 6.

59. Weller J.I., Seroussi E., Ron M. Estimation of the number of genetic markers required for individual animal identification accounting for genotyping errors. *Animal Genetics* 2006; 37:387-389.

60. Yue P., Melamud E., Moult J. SNPs3D: Candidate gene and SNP selection for association studies. *BMC Bioinformatics* 2006; 7.

61. Teare M.D., Barrett J.H. Genetic epidemiology 2 - Genetic linkage studies. *Lancet* 2005; 366:1036-1044.

62. Sung Y.J., Wijsman E.M. Accounting for epistasis in linkage analysis of general pedigrees. *Human Heredity* 2007; 63:144-152.

63. Luo Y.Q., Lin S.L., Irwin M.E. Two-locus modeling of asthma in a Hutterite pedigree via Markov chain Monte Carlo. *Genetic Epidemiology* 2001; 21:S24-S29.

64. Liang K.Y., Chiu Y.F., Beaty T.H., Wjst M. Multipoint analysis using affected sib pairs: Incorporating linkage evidence from unlinked regions. *Genetic Epidemiology* 2001; 21:105-122.

65. Sham P.C., Zhao J.H., Cherny S.S., Hewitt J.K. Variance-components QTL linkage analysis of selected and non-normal samples: Conditioning on trait values. *Genetic Epidemiology* 2000; 19:S22-S28.

66. David Pearce, Giambattista Cantisani, and Aarno Laihonen. Changes in fertility and family sizes in Europe. 23-40. 1999. Office for National Statistics. Population Trends Spring 1999. 28-11-0008.
Ref Type: Report

67. Newton-Cheh C., Hirschhorn J.N. Genetic association studies of complex traits: design and analysis issues. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 2005; 573:54-69.

68. Saxena R., Voight B.F., Lyssenko V. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; 316:1331-1336.

69. Sunyaev S., Ramensky V., Koch I., Lathe W., Kondrashov A.S., Bork P. Prediction of deleterious human alleles. *Human Molecular Genetics* 2001; 10:591-597.

70. Bengert P., Dandekar T. A software tool-box for analysis of regulatory RNA elements. *Nucleic Acids Research* 2003; 31:3441-3445.

71. Lisitsky I., Klaff P., Schuster G. Blocking polyadenylation of mRNA in the chloroplast inhibits its degradation. *Plant Journal* 1997; 12:1173-1178.

72. Asthana S., Roytberg M., Stamatoyannopoulos J., Sunyaev S. Analysis of sequence conservation at nucleotide resolution. *Plos Computational Biology* 2007; 3:2559-2568.

73. Wang H.S., Thomas D.C., Pe'er I., Stram D.O. Optimal two-stage genotyping designs for genome-wide association scans. *Genetic Epidemiology* 2006; 30:356-368.

74. Ng P.C., Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* 2003; 31:3812-3814.

75. Ramensky V., Bork P., Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research* 2002; 30:3894-3900.

76. Reumers J., Maurer-Stroh S., Schymkowitz J., Rousseau F. SNPeffect v2.0: a new step in investigating the molecular phenotypic effects of human non-synonymous SNPs. *Bioinformatics* 2006; 22:2183-2185.

77. Brodsky L.I., Ivanov V.V., Kalaydzidis Y.L. et al. Genebee-Net - Internet-Based Server for Analyzing Biopolymers Structure. *Biochemistry-Moscow* 1995; 60:923-928.

78. Nalla V.K., Rogan P.K. Automated splicing mutation analysis by information theory. *Human Mutation* 2005; 25:334-342.

79. Conde L., Vaquerizas J.M., Dopazo H. et al. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Research* 2006; 34:W621-W625.

80. Wei Q., Jacobs D.R., Schreiner P.J., Siscovick D.S., Steffes M.W., Fornage M. Patterns of association between PPAR gamma genetic variation and indices of adiposity and insulin action in African-Americans and whites: the CARDIA Study. *Journal of Molecular Medicine-Jmm* 2006; 84:955-965.

81. Vaessen N., Janssen J.A., Heutink P. et al. Association between genetic variation in the gene for insulin-like growth factor-I and low birthweight. *Lancet* 2002; 359:1036-1037.

82. Frayling T.M., Hattersley A.T., McCarthy A. et al. A putative functional polymorphism in the IGF-I gene - Association studies with type 2 diabetes, adult height, glucose tolerance, and fetal growth in UK populations. *Diabetes* 2002; 51:2313-2316.

83. Rietveld I., Janssen J.A.M.J., Hofman A., Pols H.A.P., van Duijn C.M., Lamberts S.W.J. A polymorphism in the IGF-I gene influences the age-related decline in circulating total IGF-I levels. *European Journal of Endocrinology* 2003; 148:171-175.

84. Jellema A., Mensink R.P., Kromhout D., Saris W.H.M., Feskens E.J.M. Metabolic risk markers in an overweight and normal weight population with oversampling of carriers of the IRS-1 972Arg-variant. *Atherosclerosis* 2003; 171:75-81.

85. Florez J.C., Sjogren M., Burtt N. et al. Association testing in 9,000 people fails to confirm the association of the insulin receptor substrate-1 G972R polymorphism with type 2 diabetes. *Diabetes* 2004; 53:3313-3318.

86. van Dam R.M., Hoebee B., Seidell J.C., Schaap M.M., Blaak E.E., Feskens E.J.M. The insulin receptor substrate-1 Gly972Arg polymorphism is not associated with Type 2 diabetes mellitus in two population-based studies. *Diabetic Medicine* 2004; 21:752-758.

87. Pritchard J.K. Are rare variants responsible for susceptibility to complex diseases? *American Journal of Human Genetics* 2001; 69:124-137.

88. Johnson G.C.L., Esposito L., Barratt B.J. et al. Haplotype tagging for the identification of common disease genes. *Nature Genetics* 2001; 29:233-237.

89. Gorlov I.P., Gorlova O.Y., Sunyaev S.R., Spitz M.R., Amos C.I. Shifting paradigm of association studies: Value of rare single nucleotide polymorphisms. *Genetic Epidemiology* 2007; 31:608.

90. H.Cordell, D.Clayton Genetic association studies. *The Lancet* 2005; 366,1121-1131.

91. Stephens J.C., Rogers J., Ruano G. Theoretical Underpinning of the Single-Molecule-Dilution (Smd) Method of Direct Haplotype Resolution. *American Journal of Human Genetics* 1990; 46:1149-1155.

92. Douglas J.A., Boehnke M., Gillanders E., Trent J.A., Gruber S.B. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nature Genetics* 2001; 28:361-364.

93. Niu T.H., Qin Z.H.S., Xu X.P., Liu J.S. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *American Journal of Human Genetics* 2002; 70:157-169.

94. Clark A.G. Inference of Haplotypes from Pcr-Amplified Samples of Diploid Populations. *Molecular Biology and Evolution* 1990; 7:111-122.

95. Kulle B., Frigessi A., Edvardsen H., Kristensen V., Wojnowski L. Accounting for haplotype phase uncertainty in linkage disequilibrium estimation. *Genetic Epidemiology* 2008; 32:168-178.

96. Lu X., Niu T.H., Liu J.S. Haplotype information and linkage disequilibrium mapping for single nucleotide polymorphisms. *Genome Research* 2003; 13:2112-2117.

97. Jorgenson E., Witte J.S. Coverage and power in genomewide association studies. *American Journal of Human Genetics* 2006; 78:884-888.

98. Ding K., Kullo I.J. Methods for the selection of tagging SNPs: a comparison of tagging efficiency and performance. *European Journal of Human Genetics* 2007; 15:228-236.

99. Halldorsson B.V., Istrail S., De la Vega F.M. Optimal selection of SNP markers for disease association studies. *Human Heredity* 2004; 58:190-202.

100. Ding K.Y., Zhou K., Zhang J., Knight J., Zhang X.G., Shen Y. The effect of haplotype-block definitions on inference of haplotype-block structure and htSNPs selection. *Molecular Biology and Evolution* 2005; 22:148-159.

101. BR Korf Human genetics and genomics. *Wiley-Blackwell,* 2006.

102. Halperin E., Karp R.M. The minimum-entropy set cover problem. *Automata , Languages and Programming, Proceedings* 2004; 3142:733-744.

103. Zhang K., Qin Z.H., Chen T., Liu J.S., Waterman M.S., Sun F.Z. HapBlock: haplotype block partitioning and tag SNP selection software using a set of dynamic programming algorithms. *Bioinformatics* 2005; 21:131-134.

104. Ellen L Goode, Brooke L Fridley, Zhifu Sun, Elizabeth J Atkinson, Alex S Nord, Shannon K McDonnell, Gail P Jarvik, Mariza de Andrade, and Susan L Slager. Comparison of tagging single-nucleotide polymorphism methods in association analyses. BMC Proceedings 1(Suppl 1): S6. 18-12-2007.
Ref Type: Conference Proceeding

105. Barrett J.C., Fry B., Maller J., Daly M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; 21:263-265.

106. Colin M.Clarke. NanoSNP: A high throughput SNP genotyping system based on semiconductor quantum dots. 2004. IBSAT Institute of Bioscience and Analytical Technology, DASI Department of Analytical Science and Informatics.
Ref Type: Thesis/Dissertation

107. Stephens R.H., McElduff P., Heald A.H. et al. Polymorphisms in IGF-binding protein 1 are associated with impaired renal function in type 2 diabetes. *Diabetes* 2005; 54:3547-3553.

108. Conde L.V.J.M.F.-C.C.O.M.&.D.J. PupasView: a visual tool for selecting suitable SNPs, with putative pathological effect in genes, for genotyping purposes. *Nucl Acids Res* 2005; 33:501-505.

109. Conde L.V.J.M.S.J.A.-S.F.R.-L.S.R.M.&.D.J. PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucl Acids Res* 2004; 33:242-248.

110. Cartegni L C.S.K.AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002; 4:285-298.

111. B R Graveley Sorting out the complexity of SR protein functions. *RNA* 2000; 6:1197-1211.

112. Cartegni L W.J.Z.Z.Z.M.K.AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; 31:3568-3571.

113. Patel N.A., Kaneko S., Apostolatos H.S. et al. Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase C beta II alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *Journal of Biological Chemistry* 2005; 280:14302-14309.

114. Patel N.A., Chalfant C.E., Watson J.E. et al. Insulin regulates alternative splicing of protein kinase C beta II through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor, SRp40, in skeletal muscle cells. *Journal of Biological Chemistry* 2001; 276:22648-22654.

115. Smith P.J., Spurrell E.L., Coakley J. et al. An exonic splicing enhancer in human IGF-I pre-mRNA mediates recognition of alternative exon 5 by the serine-arginine protein splicing factor-2/alternative splicing factor. *Endocrinology* 2002; 143:146-154.

116. Cartegni L., Krainer A.R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature Genetics* 2002; 30:377-384.

117. Rininsland F., Johnson T.R., Chernicky C.L. et al. Suppression of insulin-like growth factor type I receptor by a triple-helix strategy inhibits IGF-I transcription and tumorigenic potential of rat C6 glioblastoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 1997; 94:5854-5859.

118. Goni JR d.l.C.X.O.M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res* 2004; 32:354-360.

119. Gaddis SS W.Q.T.H.D.J.W.E.M.M.V.KM. A web-based search engine for triplex-forming oligonucleotide target sequences. *Oligonucleotides* 2006; 16:196-201.

120. Berg U., Gustafsson T., Sundberg C.J. et al. Local changes in the insulin-like growth factor system in human skeletal muscle assessed by microdialysis and arterio-venous differences technique. *Growth Hormone & Igf Research* 2006; 16:217-223.

121. Waters P.J., Parniak M.A., Akerman B.R., Scriver C.R. Characterization of phenylketonuria missense substitutions, distant from the phenylalanine

hydroxylase active site, illustrates a paradigm for mechanism and potential modulation of phenotype, (vol 69, pg 101, 2000). *Molecular Genetics and Metabolism* 2001; 72:89.

122. Conticello S.G., Pilpel Y., Glusman G., Fainzilber M. Position-specific codon conservation in hypervariable gene families. *Trends in Genetics* 2000; 16:57-59.

123. Kotlar D., Lavner Y. The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC Genomics* 2006; 7.

124. Kanaya S., Yamada Y., Kinouchi M., Kudo Y., Ikemura T. Codon usage and tRNA genes in eukaryotes: Correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *Journal of Molecular Evolution* 2001; 53:290-298.

125. Kleppe K., Ohtsuka E., Kleppe R., Molineux I., Khorana H.G. Studies on Polynucleotides .96. Repair Replication of Short Synthetic Dnas As Catalyzed by Dna Polymerases. *Journal of Molecular Biology* 1971; 56:341-&.

126. Mullis K.B. The Unusual Origin of the Polymerase Chain-Reaction. *Scientific American* 1990; 262:56-&.

127. Mikael Kubista, José Manuel Andrade, Martin Bengtsson et al. The real-time polymerase chain reaction. *Molecular Aspects of Medicine* 2006; 27:95-125.

128. Andy Vierstraete. Principle of the PCR. Andy Vierstraete Homepage . 8-11-1999. 27-9-2006.
Ref Type: Electronic Citation

129. M.J.McPherson, Graham R.Taylor, P.Quirke PCR: A Practical Approach. *Oxford University Press,* 1991: 1.

130. Norman Arnheim, Henry Erlich Polymerase Chain Reaction Strategy. *Annual Review of Biochemistry* 1992; 61:131-156.

131. John D.Pfeifer, Daniel A.Arber Molecular Genetic Testing in Surgical Pathology. *Lippincott Williams & Wilkins,* 2006: 86.

132. Dieffenbach C.W., Lowe T.M.J., Dveksler G.S. General Concepts for Pcr Primer Design. *Pcr-Methods and Applications* 1993; 3:S30-S37.

133. Sanchez J.J., Borsting C., Hallenberg C., Buchard A., Hernandez A., Morling N. Multiplex PCR and minisequencing of SNPs - a model with 35 Y chromosome SNPs. *Forensic Science International* 2003; 137:74-84.

134. Feng Y., Prohofsky E.W. Vibrational Fluctuations of Hydrogen-Bonds in A Dna Double Helix with Nonuniform Base-Pairs. *Biophysical Journal* 1990; 57:547-553.

135. Anton Yuryev PCR Primer Design. *Humana Press,* 2007: 73.

136. Ayyadevara S., Thaden J.J., Reis R.J.S. Discrimination of primer 3 '-nucleotide mismatch by Taq DNA polymerase during polymerase chain reaction. *Analytical Biochemistry* 2000; 284:11-18.

137. Lowe T., Sharefkin J., Yang S.Q., Dieffenbach C.W. A Computer-Program for Selection of Oligonucleotide Primers for Polymerase Chain Reactions. *Nucleic Acids Research* 1990; 18:1757-1761.

138. F.John Burpo. A critical review of PCR primer design algorithms and cross hybridization case study. CA 94305. 11-8-2001. Stanford University - Department of Chemical Engineering. 20-1-2009.
Ref Type: Report

139. Richard Owczarzy and Mark Behlke. Calculation of Tm for Oligonucleotide Duplexes. Integrated DNA Technologies . 2007. 16-5-0007.
Ref Type: Electronic Citation

140. Strachan T., Read A.P. Human Molecular Genetics 2 - DNA structure and gene expression. *BIOS Scientific Publishers Ltd,* 1999.

141. Panjkovich A., Melo F. Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics* 2005; 21:711-722.

142. Vallone P.M., Butler J.M. AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* 2004; 37:226-231.

143. Kamel Abd-Elsalam Bioinformatic tools and guideline for PCR primer design. *African Journal of Biotechnology* 2008; 2:91-95.

144. Cariello N.F., Thilly W.G., Swenberg J.A., Skopek T.R. Deletion Mutagenesis During Polymerase Chain-Reaction - Dependence on Dna-Polymerase. *Gene* 1991; 99:105-108.

145. Viguera E., Canceill D., Ehrlich S.D. In vitro replication slippage by DNA polymerases from thermophilic organisms. *Journal of Molecular Biology* 2001; 312:323-333.

146. Kim S., Misra A. SNP genotyping: Technologies and biomedical applications. *Annual Review of Biomedical Engineering* 2007; 9:289-320.

147. Hoheisel J.D. Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics* 2006; 7:200-210.

148. Epstein J.R., Biran I., Walt D.R. Fluorescence-based nucleic acid detection and microarrays. *Analytica Chimica Acta* 2002; 469:3-36.

149. Clavijo A., Hole K., Li M.Y., Collignon B. Simultaneous detection of antibodies to foot-and-mouth disease non-structural proteins 3ABC, 3D, 3A and 3B by a multiplexed Luminex assay to differentiate infected from vaccinated cattle. *Vaccine* 2006; 24:1693-1704.

150. Dunbar S., Godbout R., Newkirk H., Hetzel J. Microsphere suspension array technology for SNP detection in cattle. *Ieee Engineering in Medicine and Biology Magazine* 2003; 22:158-+.

151. Syvanen A.C. Toward genome-wide SNP genotyping. *Nature Genetics* 2005; 37:S5-S10.

152. Letowski J., Brousseau R., Masson L. Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *Journal of Microbiological Methods* 2004; 57:269-278.

153. Armstrong B., Stewart M., Mazumder A. Suspension arrays for high throughput, multiplexed single nucleotide polymorphism genotyping. *Cytometry* 2000; 40:102-108.

154. Xiangfeng Wang, Lei Li, Viktor Stolc, Waraporn Tongprasit, Chen Chen, Jun Wang, Songgang Li, and Xing Wang Deng. Analysis of oligo hybridization properties by high-resolution tiling microarrays in rice. 3-5-2007. Yale. 1-1-2008.
Ref Type: Report

155. Steve Rozen, Helen Skaletsky Bioinformatics Methods and Protocols. *Humana Press,* 1999: 365.

156. Breslauer K.J., Frank R., Blocker H., Marky L.A. Predicting DNA Duplex Stability from the Base Sequence. *Proceedings of the National Academy of Sciences* 1986; 83:3746-3750.

157. Leber M., Kaderali L., Schonhuth A., Schrader R. A fractional programming approach to efficient DNA melting temperature calculation. *Bioinformatics* 2005; 21:2375-2382.

158. Schoske R., Vallone P.M., Ruitberg C.M., Butler J.M. Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Analytical and Bioanalytical Chemistry* 2003; 375:333-343.

159. Cheryl D.Harrison, Chris Phillips, Claus Børsting et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *ELECTROPHORESIS* 2006; 27:1713-1724.

160. Butler J.M., Ruitberg C.M., Vallone P.M. Capillary electrophoresis as a tool for optimization of multiplex PCR reactions. *Fresenius Journal of Analytical Chemistry* 2001; 369:200-205.

161. Inagaki S., Yamamoto Y., Doi Y. et al. A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Science International* 2004; 144:45-57.

162. Henegariu O., Heerema N.A., Dlouhy S.R., Vance G.H., Vogt P.H. Multiplex PCR: Critical parameters and step-by-step protocol. *Biotechniques* 1997; 23:504-511.

163. Borsting C., Sanchez J.J., Morling N. Multiplex PCR, amplicon size and hybridization efficiency on the NanoChip electronic microarray. *International Journal of Legal Medicine* 2004; 118:75-82.

164. Oda S., Ueyama H., Nishida Y., Tanabe S., Yamade S. Analysis of L-cone/M-cone visual pigment gene arrays in females by long-range PCR. *Vision Research* 2003; 43:489-495.

165. Lane S, Evermann J., Loge F., Call D.R. Amplicon secondary structure prevents target hybridization to oligonucleotide microarrays. *Biosensors and Bioelectronics* 2004; 20:728-735.

166. Liu W.T., Guo H., Wu J.H. Effects of Target Length on the Hybridization Efficiency and Specificity of rRNA-Based Oligonucleotide Microarrays. *Applied and Environmental Microbiology* 2007; 73:73-82.

167. Shchepinov M.S., CaseGreen S.C., Southern E.M. Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. *Nucleic Acids Research* 1997; 25:1155-1161.

168. Southern E., Mir K., Shchepinov M. Molecular interactions on microarrays. *Nat Genet* 1999; 21:5-9.

169. Dunbar S.A., Jacobson J.W. Application of the Luminex LabMAP in rapid screening for mutations in the cystic fibrosis transmembrane conductance regulator gene: A pilot study. *Clinical Chemistry* 2000; 46:1498-1500.

170. Eckert K.A, Kunkel T.A DNA polymerase fidelity and the polymerase chain reaction. *Genome Res* 1991; 1:17-24.

171. Anbar M., Bracha R., Nuchamowitz Y., Li Y., Florentin A., Mirelman D. Involvement of a short interspersed element in epigenetic transcriptional silencing of the amoebapore gene in Entamoeba histolytica. *Eukaryotic Cell* 2005; 4:1775-1784.

172. Smith-Sorensen B., Lind G.E., Skotheim R.I. et al. Frequent promoter hypermethylation of the O6-Methylguanine-DNA Methyltransferase (MGMT) gene in testicular cancer. *Oncogene* 2002; 21:8878-8884.

173. Jiang M.H., Fei J., Lan M.S. et al. Hypermethylation of hepatic Gck promoter in ageing rats contributes to diabetogenic potential. *Diabetologia* 2008; 51:1525-1533.

174. Peter A.Davies, George Gray PCR Mutation Detection Protocols. 2002: 51.

175. Ruslan Kalendar. FastPCR User's Manual (beta version). University of Helsinki, Institute of Biotechnology . 2009. 30-1-0009.
Ref Type: Electronic Citation

176. Peter M.Vallone, Rebecca S.Just, Michael D.Coble, John M.Butler, Thomas J.Parsons A multiplex allele-specific primer extension assay for forensically

informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med* 2004; 118:147-157.

177. Jeffrey A.Manthey. mFold, Delta G, and Melting Temperature: What Does it Mean? IDT Integrated DNA Technologies Technical Support . 2005. 29-1-0009.
Ref Type: Electronic Citation

178. Mara R.Diaz, Jack W.Fell High-Throughput Detection of Pathogenic Yeasts of the Genus *Trichosporon*. *J Clin Microbiol* 2004; 42:3696-3706.

179. Hyone-Myong Eun Enzymology Primer for Recombinant DNA Technology. *Academic Press,* 1996: 407.

180. Markoulatos P., Siafakas N., Moncany M. Multiplex polymerase chain reaction: A practical approach. *Journal of Clinical Laboratory Analysis* 2002; 16:47-51.

181. Mutter G.L., Boynton K.A. PCR Bias in Amplification of Androgen Receptor Alleles, A Trinucleotide Repeat Marker Used in Clonality Studies. *Nucleic Acids Research* 1995; 23:1411-1418.

182. Abu Al-Soud W., Radstrom P. Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *Journal of Clinical Microbiology* 2000; 38:4463-4470.

183. Hite J.M., Eckert K.A., Cheng K.C. Factors affecting fidelity of DNA synthesis during PCR amplification of d(C-A)(n)center dot d(G-T)(n) microsatellite repeats. *Nucleic Acids Research* 1996; 24:2429-2434.

184. Chou Q. Minimizing Deletion Mutagenesis Artifact During Taq Dna-Polymerase Pcr by Escherichia-Coli Ssb. *Nucleic Acids Research* 1992; 20:4371.

185. Kellogg D.E., Rybalkin I., Chen S. et al. *Taq*start Antibody(Tm) - Hot Start Pcr Facilitated by A Neutralizing Neutralizing Monoclonal-Antibody Directed Against Taq Dna-Polymerase. *Biotechniques* 1994; 16:1134-1137.

186. Clontech Laboratories. Effective Hot Start PCR: TaqStart® Antibody for fast, convenient hot start. Clontechniques . 2008. Clontech Laboratories. 10-1-2009.
Ref Type: Catalog

187. Applied Biosystems. AmpliTaq Gold® PCR Master Mix. [P/N 4317621 Rev. B]. 4-4-2001. 2-2-2008.
Ref Type: Catalog

188. USB Corporation. HotStart-IT®: A Novel Hot Start PCR Method Based on Primer Sequestration. 11-1-2008. 12-6-2006.
Ref Type: Patent

189. Barnes W.M. The Fidelity of Taq Polymerase Catalyzing Pcr Is Improved by An N-Terminal Deletion. *Gene* 1992; 112:29-35.

190. Cariello N.F., Swenberg J.A., Skopek T.R. Fidelity of Thermococcus-Litoralis Dna-Polymerase (Vent) in Pcr Determined by Denaturing Gradient Gel-Electrophoresis. *Nucleic Acids Research* 1991; 19:4193-4198.

191. Cha R.S., Thilly W.G. Specificity, Efficiency, and Fidelity of Pcr. *Pcr-Methods and Applications* 1993; 3:S18-S29.

192. Eckert K.A., Kunkel T.A. High Fidelity Dna-Synthesis by the Thermus-Aquaticus Dna-Polymerase. *Nucleic Acids Research* 1990; 18:3739-3744.

193. Cline J., Braman J.C., Hogrefe H.H. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research* 1996; 24:3546-3551.

194. Mullan B., Sheehy P., Shanahan F., Fanning L. Do Taq-generated RT-PCR products from RNA viruses accurately reflect viral genetic heterogeneity? *Journal of Viral Hepatitis* 2004; 11:108-114.

195. Ghadessy F.J., Ong J.L., Holliger P. Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* 2001; 98:4552-4557.

196. Pavlov A.R., Pavlova N.V., Kozyavkin S.A., Slesarev A.I. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends in Biotechnology* 2004; 22:253-260.

197. Rasmussen T., Poulsen T.S., Honore L., Johnsen H.E. Quantitation of minimal residual disease in multiple myeloma using an allele-specific real-time PCR assay. *Experimental Hematology* 2000; 28:1039-1045.

198. Bergallo M., Costa C., Baro S. et al. Multiplex-nested RT-PCR to evaluate latent and lytic Epstein Barr virus gene expression. *Journal of Biotechnology* 2007; 128:462-476.

199. Stram Y., Kuznetzova L., Guini M. et al. Detection and quantitation of Akabane and Aino viruses by multiplex real-time reverse-transcriptase PCR. *Journal of Virological Methods* 2004; 116:147-154.

200. Shivapurkar N., Stastny V., Suzuki M. et al. Application of a methylation gene panel by quantitative PCR for lung cancers. *Cancer Letters* 2007; 247:56-71.

201. Choi B.O., Kim J., Lee K.L., Yu J.S., Hwang J.H., Chung K.W. Rapid diagnosis of CMT1A duplications and HNPP deletions by multiplex microsatellite PCR. *Molecules and Cells* 2007; 23:39-48.

202. Poe B.G., Navratil M., Arriaga E.A. Absolute quantitation of a heteroplasmic mitochondrial DNA deletion using a multiplex three-primer real-time PCR assay. *Analytical Biochemistry* 2007; 362:193-200.

203. Szilvasi A., Andrikovics H., Kalmar L., Bors A., Tordai A. Asymmetric PCR increases efficiency of melting peak analysis on the LightCycler. *Clinical Biochemistry* 2005; 38:727-730.

204. Skrzypski M. Quantitative reverse transcriptase real-time polymerase chain reaction (qRT-PCR) in translational oncology: Lung cancer perspective. *Lung Cancer* 2008; 59:147-154.

205. Priya K., Madhavan H.N., Malathi J. Use of uniplex polymerase chain reaction & evaluation of multiplex PCR in the rapid diagnosis of viral retinitis. *Indian Journal of Medical Research* 2003; 117:205-210.

206. Deepak S.A., Kottapalli K.R., Rakwal R. et al. Real-time PCR: Revolutionizing detection and expression analysis of genes. *Current Genomics* 2007; 8:234-251.

207. Gerard C.J., Olsson K., Ramanathan R., Reading C., Hanania E.G. Improved quantitation of minimal residual disease in multiple myeloma using real-time polymerase chain reaction and Plasmid-DNA complementarity determining region III standards. *Cancer Research* 1998; 58:3957-3964.

208. Wilhelm J., Pingoud A. Real-time polymerase chain reaction. *Chembiochem* 2003; 4:1120-1128.

209. Joanne Traeger-Synodinos Real-time PCR for prenatal and preimplantation genetic diagnosis of monogenic diseases. *Molecular Aspects of Medicine* 2006; 27:176-191.

210. Gudnason H., Dufva M., Bang D.D., Wolff A. Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic Acids Research* 2007; 35.

211. Y.M.Dennis Lo Clinical Applications of PCR. *Humana Press,* 2006: 19.

212. Price E.P., Smith H., Huygens F., Giffard P.M. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of Campylobacter jejuni. *Applied and Environmental Microbiology* 2007; 73:3431-3436.

213. Elnifro E.M., Ashshi A.M., Cooper R.J., Klapper P.E. Multiplex PCR: Optimization and application in diagnostic virology. *Clinical Microbiology Reviews* 2000; 13:559-+.

214. Bose B., Chugh D.A., Kala M., Acharya S.K., Khanna N., Sinha S. Characterization and molecular modeling of a highly stable anti-Hepatitis B surface antigen scFv. *Molecular Immunology* 2003; 40:617-631.

215. Qiagen QIAGEN Multiplex PCR Handbook. *Qiagen,* 2008.

216. Fanali S. Identification of chiral drug isomers by capillary electrophoresis. *Journal of Chromatography A* 1996; 735:77-121.

217. Tsai C.H., Fang C., Liu J.T., Lin C.H. Stacking and low-temperature technique in nonaqueous capillary electrophoresis for the analysis of 3,4-methylenedioxymethamphetamine. *ELECTROPHORESIS* 2004; 25:1601-1606.

218. Chen Y.H., Chung Y.L., Lin C.H. Ultra-low-temperature non-aqueous capillary electrophoretic separation-77 K fluorescence spectroscopic detection for the on-line identification of photo-converted analytes of trans-resveratrol. *Journal of Chromatography A* 2002; 943:287-294.

219. Chamberlain J.S., Gibbs R.A., Ranier J.E., Nguyen P.N., Caskey C.T. Deletion Screening of the Duchenne Muscular-Dystrophy Locus Via Multiplex Dna Amplification. *Nucleic Acids Research* 1988; 16:11141-11156.

220. Berg J M., Tymoczko J L., Stryer L. Biochemistry. W. H. Freeman and Company, 2002.

221. Hunt M. Microbiology and Immunology online (Real-time PCR tutorial). 2008.

222. Pfaffl M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 2001; 29.

223. Jo Vandesompele, Katleen De Preter, Filip Pattyn et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002; 3.

224. Fan W., Hamilton T., Webster-Sesay S., Nikolich M.P., Lindler L.E. Multiplex real-time SYBR Green IPCR assay for detection of tetracycline efflux genes of Gram-negative bacteria. *Molecular and Cellular Probes* 2007; 21:245-256.

225. Varga A., James D. Detection and differentiation of Plum pox virus using real-time multiplex PCR with SYBR Green and melting curve analysis: a rapid method for strain typing. *Journal of Virological Methods* 2005; 123:213-220.

226. Rachlin J., Ding C.M., Cantor C., Kasif S. Computational tradeoffs in multiplex PCR assay design for SNP genotyping. *BMC Genomics* 2005; 6.

227. Polysciences Inc. Microsphere Selection. Technical datasheet 778. 15-9-2008. POlysciences Inc.
Ref Type: Pamphlet

228. Spherotech Inc. Spherotech Products Page. Spherotech website . 2008. 14-3-2008.
Ref Type: Electronic Citation

229. Luminex Corportaion. Luminex Products Page - Building a Custom Assay. Luminex Corporation Website . 2009.
Ref Type: Electronic Citation

230. Bangs Laboratories, Inc. Bangs Laboratories, Inc.- Products and Ordering Page. Bands Laboratories, Inc.Website . 2009. 3-4-2007.
Ref Type: Electronic Citation

231. Kellar K.L., Iannone M.A. Multiplexed microsphere-based flow cytometric assays. *Experimental Hematology* 2002; 30:1227-1237.

232. Jin R.C., Cao Y.C., Thaxton C.S., Mirkin C.A. Glass-bead-based parallel detection of DNA using composite Raman labels. *Small* 2006; 2:375-380.

233. Kneipp K., Kneipp H., Manoharan R., Itzkan I., Dasari R.R., Feld M.S. Near-infrared surface-enhanced Raman scattering can detect single molecules and observe 'hot' vibrational transitions. *Journal of Raman Spectroscopy* 1998; 29:743-747.

234. Mirkin C. Nanoparticles with raman spectroscopic fingerprints for DNA and RNA detection. *Abstracts of Papers of the American Chemical Society* 2003; 225:U963-U964.

235. Doering W.E., Nie S.M. Spectroscopic tags using dye-embedded nanoparticies and surface-enhanced Raman scattering. *Analytical Chemistry* 2003; 75:6171-6176.

236. Moran E.J., Sarshar S., Cargill J.F. et al. Radio-Frequency Tag Encoded Combinatorial Library Method for the Discovery of Tripeptide-Substituted Cinnamic Acid Inhibitors of the Protein-Tyrosine-Phosphatase Ptp1B. *Journal of the American Chemical Society* 1995; 117:10787-10788.

237. Nolan J.P., Mandy F. Multiplexed and microparticle-based analyses: Quantitative tools for the large-scale analysis of biological systems. *Cytometry Part A* 2006; 69A:318-325.

238. Chan W.C.W., Nie S.M. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* 1998; 281:2016-2018.

239. Ekimov A.I., Efros Al.L., Onushchenko A.A. Quantum size effect in semiconductor microcrystals. *Solid State Communications* 1985; 56:921-924.

240. Han M.Y., Gao X.H., Su J.Z., Nie S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nature Biotechnology* 2001; 19:631-635.

241. Crystalplex. Quantum dot encoded microsphere scheme. CRystalplex Website . 2007. 12-11-2007.
Ref Type: Electronic Citation

242. Morgan E., Varro R., Sepulveda H. et al. Cytometric bead array: a multiplexed assay platform with applications in various areas of biology. *Clinical Immunology* 2004; 110:252-266.

243. Tarnok A., Hambsch J., Chen R., Varro R. Cytometric bead array to measure six cytokines in twenty-five microliters of serum. *Clinical Chemistry* 2003; 49:1000-1002.

244. Resch-Genger U., Grabolle M., Cavaliere-Jaricot S., Nitschke R., Nann T. Quantum dots versus organic dyes as fluorescent labels. *Nature Methods* 2008; 5:763-775.

245. Jamieson T., Bakhshi R., Petrova D., Pocock R., Imani M., Seifalian A.M. Biological applications of quantum dots. *Biomaterials* 2007; 28:4717-4732.

246. Csako G. Present and future of rapid and/or high-throughput methods for nucleic acid testing. *Clinica Chimica Acta* 2006; 363:6-31.

247. Z Tsuchihashi, N C Dracopoli Progress in high throughput SNP genotyping methods. *The Pharmacogenomics Journal* 2002; 2:103-110.

248. Rao R.S., Visuri S.R., McBride M.T., Albala J.S., Matthews D.L., Coleman M.A. Comparison of multiplexed techniques for detection of bacterial and viral proteins. *Journal of Proteome Research* 2004; 3:736-742.

249. Reeve L., Rew D.A. New technology in the analytical cell sciences: the laser scanning cytometer. *European Journal of Surgical Oncology* 1997; 23:445-450.

250. Darzynkiewicz Z., Bedner E., Li X., Gorczyca W., Melamed M.R. Laser-scanning cytometry: A new instrumentation with many applications. *Experimental Cell Research* 1999; 249:1-12.

251. Wilson M.H., Grant P.J., Kain K., Warner D.P., Wild C.P. Association between the risk of coronary artery disease in South Asians and a deletion polymorphism in glutathione S-transferase M1. *Biomarkers* 2003; 8:43-50.

252. Schmitt M., Bravo I.G., Snijders P.J.F., Gissmann L., Pawlita M., Waterboer T. Bead-based multiplex genotyping of human papillomaviruses. *Journal of Clinical Microbiology* 2006; 44:504-512.

253. Flagella M., Bui S., Zheng Z. et al. A multiplex branched DNA assay for parallel quantitative gene expression profiling. *Analytical Biochemistry* 2006; 352:50-60.

254. Li Z.P., Kambara H. Single nucleotide polymorphism analysis based on minisequencing coupled with a fluorescence microsphere technology. *Journal of Nanoscience and Nanotechnology* 2005; 5:1256-1260.

255. Deregt D., Gilbert S.A., Dudas S. et al. Multiplex DNA suspension microarray for simultaneous detection and differentiation of classical swine fever virus and other pestiviruses. *Journal of Virological Methods* 2006; 136:17-23.

256. Lowe D., Hathaway M., Briggs D. The high-dose hook effect in the detection and monitoring of HLA specific antibody by Luminex assay. *International Journal of Immunogenetics* 2007; 34:288.

257. Smith J.D., Rose M.L. Development of a Luminex based method to detect C4d binding by HLA antibodies. *International Journal of Immunogenetics* 2007; 34:290.

258. Yurkovetsky Z.R., Kirkwood J.M., Edington H.D. et al. Multiplex analysis of serum cytokines in melanoma patients treated with interferon-alpha 2b. *Clinical Cancer Research* 2007; 13:2422-2428.

259. Dehqanzada Z.A., Storrer C.E., Hueman M.T. et al. Assessing serum cytokine profiles in breast cancer patients receiving a HER2/neu vaccine using Luminex (R) technology. *Annals of Surgical Oncology* 2005; 12:S47-S48.

260. Fulton R.J., Mcdade R.L., Smith P.L., Kienker L.J., Kettman J.R. Advanced multiplexed analysis with the FlowMetrix(TM) system. *Clinical Chemistry* 1997; 43:1749-1756.

261. Becton Dickinson. BD FACSAria Flow Cytometer: The first high-speed sorter with a fixed-alignment cuvette flow cell from BD Biosciences. 11-5-2007.
Ref Type: Catalog

262. Galbraith D.W. Cytometry and plant sciences: a personal retrospective. *Cytometry Part A* 2004; 58A:37-44.

263. Doyen T.Nguyen, Lawrence W.Diamond, Raul C.Braylan Flow Cytometry in Hematopathology: A Visual Approach to Data Analysis and Interpretation. *Humana Press,* 2007: 50-64.

264. James V.Watson Introduction to Flow Cytometry. *Cambridge University Press,* 1991.

265. Chapman G.V. Instrumentation for flow cytometry. *Journal of Immunological Methods* 2000; 243:3-12.

266. SpheroTech. Measuring molecules of equivalent fluorochrome (MEF) using Sphero Rainbow and Ultra Rainbow Calibration Particles. 4-2-2008. 2-5-2007.
Ref Type: Pamphlet

267. Becman Coulter COULTER® EPICS® XL™Flow Cytometer. 2000.

268. De Novo. FCS Express. [3.00.0707]. 13-2-2009.
Ref Type: Computer Program

269. National Science Foundation. Counting Cells with a Hemocytometer - Protocol. 18-10-2006.
Ref Type: Report

270. James L.Weaver Introduction to Flow Cytometry. *Methods* 2000; 21:199-201.

271. Ademtech. MasterBeads Carboxylic Acid (0215). version 1.2. 18-6-2007. Adamtech. 3-5-2007.
Ref Type: Pamphlet

272. Newkirk H.L., Rogan P.K., Miralles M., Knoll J.H.M. Determination of genomic copy number with quantitative microsphere hybridization. *Human Mutation* 2006; 27:376-386.

273. Nikiforov, Theo and Knapp, Michael R. Method for immobilizing nucleic acid molecules . 08/341148[5610287]. 3-11-1997. United States.
Ref Type: Patent

274. Dunbar S.A. Applications of Luminex (R) xMAP (TM) technology for rapid, high-throughput multiplexed nucleic acid detection. *Clinica Chimica Acta* 2006; 363:71-82.

275. Crystalplex Corp. Crystalplex PlxBeads Product Page. Crystalplex Corp.Pittsburgh, PA. 15-11-2008. 4-12-2008.
Ref Type: Electronic Citation

276. Wood W.I., Gitschier J., Lasky L.A., Lawn R.M. Base Composition-Independent Hybridization in Tetramethylammonium Chloride - A Method for Oligonucleotide Screening of Highly Complex Gene Libraries. *Proceedings of the National Academy of Sciences of the United States of America* 1985; 82:1585-1588.

277. Das S., Brown T.M., Kellar K.L., Holloway B.P., Morrison C.J. DNA probes for the rapid identification of medically important Candida species using a multianalyte profiling system. *FEMS Immunology & Medical Microbiology* 2006; 46:244-250.

278. B.D.Hames, S.J.Higgins Gene probes 2: A Practical Approach. *Oxford University Press,* 1995: 1.

279. Puzin Y.I., Yumagulova R.K., Kraikin V.A. Radical polymerization of methyl methacrylate and styrene in the presence of ferrocene. *European Polymer Journal* 2001; 37:1801-1812.

280. Chen M., Zhou S.X., Wu L.M., Xie S.H., Chen Y. Preparation of silica-coated polystyrene hybrid spherical colloids. *Macromolecular Chemistry and Physics* 2005; 206:1896-1902.

281. Cao Y.C., Huang Z.L., Liu T.C. et al. Preparation of silica encapsulated quantum dot encoded beads for multiplex assay and its properties. *Analytical Biochemistry* 2006; 351:193-200.

282. Saito R., Kobayashi S.I., Hayashi H., Shimo T. Surface hardness and transparency of poly(methyl methacrylate)-silica coat film derived from perhydropolysilazane. *Journal of Applied Polymer Science* 2007; 104:3388-3395.

283. Barthet C., Hickey A.J., Cairns D.B., Armes S.P. Synthesis of novel polymer-silica colloidal nanocomposites via free-radical polymerization of vinyl monomers. *Advanced Materials* 1999; 11:408-410.

284. Amalvy J.I., Percy M.J., Armes S.P., Wiese H. Synthesis and characterization of novel film-forming vinyl polymer/silica colloidal nanocomposites. *Langmuir* 2001; 17:4770-4778.

285. Rao K.V.N., Stevens P.W., Hall J.G., Lyamichev V., Neri B.P., Kelso D.M. Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage reaction on microspheres. *Nucleic Acids Research* 2003; 31.

286. Deshpande A, Yolanda Valdez, John P.Nolan Current Protocols in Cytometry. 2005.

287. Becton Dickinson. QuantiBRITE PE Phycoerythrin Fluorescence Quantitation Kit. 12-3-2001.
Ref Type: Pamphlet

288. Spherotech, Inc. SPHERO Calibration Particles. 24-10-2007. Spherotech inc. 5-6-0007.
Ref Type: Pamphlet

289. A.Schwartz, G.E.Marti, R.Poon, J.W.Gratama, E.Ferna´ndez-Repollet Standardizing Flow Cytometry: A Classification System of Fluorescence Standards Used for Flow Cytometry. *Cytometry* 1998; 33:106-114.

290. Hye-Jung Yeoma, Young-Sun Hera, Moon-Ju Oha et al. Application of multiplex bead array assay for Yq microdeletion analysis in infertile males. *Molecular and Cellular Probes* 2008; 22:76-82.

291. Richard Morilla and Mariano Scolnik. Quantitative Flow Cytometry. 33[4], 267-270. 1-12-1998. 7-12-2007.
Ref Type: Conference Proceeding

292. Robert A.Hoffman Current Protocols in Cytometry. 2005.

293. Eric J.Devor and Mark A.Behlke. Strategies for Attaching Oligonucleotides to Solid Supports. 2005. Integrated DNA Technologies.
Ref Type: Pamphlet

294. Lawrie G.A., Robinson J., Corrie S., Ford K., Battersby B.J., Trau M. Multiplexed microsphere diagnostic tools in gene expression applications: factors and futures. *International Journal of Nanomedicine* 2006; 1:195-201.

295. Wittebolle L., Verstuyft K., Verstraete W., Boon N. Optimisation of the amino-carboxy coupling of oligonucleotides to beads used in liquid arrays. *Journal of Chemical Technology and Biotechnology* 2006; 81:476-480.

296. Fuja T., Hou S., Bryant P. A multiplex microsphere bead assay for comparative RNA expression analysis using flow cytometry. *Journal of Biotechnology* 2004; 108:193-205.

297. Yao G., Tan W.H. Molecular-beacon-based array for sensitive DNA analysis. *Analytical Biochemistry* 2004; 331:216-223.

298. Wetmur J.G. Dna Probes - Applications of the Principles of Nucleic-Acid Hybridization. *Critical Reviews in Biochemistry and Molecular Biology* 1991; 26:227-259.

299. Sanchez J.J., Phillips C., Borsting C. et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *ELECTROPHORESIS* 2006; 27:1713-1724.

300. Makowski G.S., Nadeau F.L., Hopfer S.M. Single tube multiplex PCR detection of 27 cystic fibrosis mutations and 4 polymorphisms using neonatal blood samples collected on guthrie cards. *Annals of Clinical and Laboratory Science* 2003; 33:243-250.

301. Harries L.W., Stubbins M.J., Forman D., Howard G.C.W., Wolf C.R. Identification of genetic polymorphisms at the glutathione S-transferase Pi locus

and association with susceptibility to bladder, testicular and prostate cancer. *Carcinogenesis* 1997; 18:641-644.

302. Salama S.A., Sierra-Torres C.H., Oh H.Y., Hamada F.A., Au W.W. A multiplex-PCR/RFLP procedure for simultaneous CYP2E1, mEH and GSTM1 genotyping. *Cancer Letters* 1999; 143:51-56.

303. Smith C.A.D., Harrison D.J. Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. *Lancet* 1997; 350:630-633.

304. Kihara M., Kihara M., Noda K. Lung-Cancer Risk of Gstm1 Null Genotype Is Dependent on the Extent of Tobacco-Smoke Exposure. *Carcinogenesis* 1994; 15:415-418.

305. Mirel D.B., Valdes A.M., Lazzeroni L.C., Reynolds R.L., Erlich H.A., Noble J.A. Association of IL4R haplotypes with type 1 diabetes. *Diabetes* 2002; 51:3336-3341.

306. Fitzgerald C., Collins M., van Duyne S., Mikoleit M., Brown T., Fields P. Multiplex, bead-based suspension array for molecular determination of common Salmonella serogroups. *Journal of Clinical Microbiology* 2007; 45:3323-3334.

307. Lins T.C.L., Nogueira L.R., Lima R.M., Gentil P., Oliveira R.J., Pereira R.W. A multiplex single-base extension protocol for genotyping Cdx2, FokI, BsmI, ApaI, and TaqI polymorphisms of the vitamin D receptor gene. *Genetics and Molecular Research* 2007; 6:316-324.

APPENDIX

1.1 Chapter 3: PCR Primer / Probe Design

Table 1: Gene name, Tagger derived bin number, tagSNP reference sequence identifier and corresponding IGF multiplex set identifiers for all primer pairs used in the final multiplex set are displayed.

Gene Name	Tagger Bin No	TagSNP Reference Sequence Identifier	IGF Multiplex Set Identifier
IGF1	1	rs972936	IGF1_1
	2	rs17882461	IGF1_2
	3	rs11111262	IGF1_3
	4	rs2033178	IGF1_4
	5	rs2946834	IGF1_5
	6	rs35767	IGF1_6
	7	rs12821878	IGF1_7
	8	rs1019731	IGF1_8
	9	rs17884646	IGF1_9
	10	rs12316064	IGF1_10
	11	rs1520220	IGF1_11
	12	rs3730204	IGF1_12
	13	rs6214	IGF1_13
IGFBP1	1	rs7454	BP1_1
	2	rs1874479	BP1_2
	3/4	rs3828998 / rs9658194	BP1_3/4
	5	rs9658238	BP1_5
IGFBP3	1	rs3793345	BP3_1
	3	rs10255707	BP3_3
	4	rs2132571	BP3_4
	6	rs11765572	BP3_6
	7	rs12540724	BP3_7
	9	rs2453839	BP3_9
	10/11	rs35751739 / rs35496550	BP3_10/11
	12	rs6670	BP3_12
	13	rs13223993	BP3_13
	14	rs10282088	BP3_14

Table 2: Chromosome and genomic base position of putative prime pair amplicons for each primer pair IGF multiplex set as determined using UCSC *in silico* PCR amplification. The chromosome and base position for IGF1_9 was derived using Ensembl.

Primer Pair Identifier	Chromosome and Base Position
IGF1_1	chr12:101349004+101349234
IGF1_2	chr12:101387933+101388175
IGF1_3	chr12:101322176+101322423
IGF1_4	chr12:101371094+101371374
IGF1_5	chr12:101311849+101312023
IGF1_6	chr12:101399601+101399774
IGF1_7	chr12:101391684+101391917
IGF1_8	chr12:101388524+101388595
IGF1_9	chr12:101,369,854-101,369,994
IGF1_10	chr12:101324870+101325068
IGF1_11	chr12:101320576+101320823
IGF1_12	chr12:101319336+101319835
IGF1_13	chr12:101317573+101317822
BP1_1	chr7:45899490+45899627
BP1_2	chr7:45898675+45898879
BP1_3/4	chr7:45895239+45895394
BP1_5	chr7:45900594+45900761
BP3_1	chr7:45924115+45924364
BP3_3	chr7:45921159+45921334
BP3_4	chr7:45928057+45928329
BP3_6	chr7:45927624+45927841
BP3_7	chr7:45923723+45923874
BP3_9	chr7:45920024+45920144
BP3_10/11	chr7:45919162+45919329
BP3_12	chr7:45918741+45918945
BP3_13	chr7:45917715+45917868
BP3_14	chr7:45916783+45917051

Table 3: Panjokovich consensus primer predicted T_As (°C) for IGF Multiplex Set. The predicted primer annealing temperature for each primer pair (both forward and reverse), mean primer pair T_A and intra-primer pair T_A difference as predicted using Primer3 predictions are displayed.

Primer Pair Identifier	Primer T_A (°C) Forward/Reverse	Mean Primer Pair T_A (°C)	Intra-primer pair T_A Difference (°C)
IGF1_1	65.76 / 64.89	65.325	0.87
IGF1_2	64.60 /65.81	65.205	1.21
IGF1_3	65.18 / 65.87	65.525	0.69
IGF1_4	64.95 / 65.35	65.15	0.40
IGF1_5	65.17 / 65.07	65.12	0.10
IGF1_6	64.95 / 65.73	65.34	0.78
IGF1_7	64.74 / 64.98	64.86	0.24
IGF1_8	65.46 / 64.83	65.145	0.63
IGF1_9	65.92 / 64.95	65.435	0.97
IGF1_10	65.87 / 65.60	65.735	0.27
IGF1_11	65.59 / 64.42	65.005	1.17
IGF1_12	64.52 / 64.43	64.475	0.09
IGF1_13	64.72 / 64.62	64.68	0.10
BP1_1	64.31 /65.14	64.725	0.83
BP1_2	64.85 / 64.62	64.735	0.23
BP1_3/4	65.64 / 65.71	65.675	0.07
BP1_5	65.41 / 65.67	65.54	0.26
BP3_1	64.82 / 65.22	65.02	0.40
BP3_3	65.77 / 65.27	65.52	0.50
BP3_4	65.56 / 65.60	65.58	0.04
BP3_6	64.64 / 65.48	65.06	0.84
BP3_7	64.83 / 65.87	65.35	1.04
BP3_9	65.47 / 65.99	65.73	0.52
BP3_10/ 11	64.53 / 65.85	65.19	1.32
BP3_12	65.66 / 65.97	65.815	0.31
BP3_13	65.63 / 65.32	65.475	0.31
BP3_14	64.82 / 65.45	65.135	0.63

Table 4: % GC for all primers (both forward and reverse), mean prime pair GC% and Intra-primer pair difference in the final IGF Multiplex set as determined by Primer3 are displayed.

Primer Pair Identifier	GC % Forward Primer	GC % Reverse Primer	Mean Primer Pair GC %	Intra-Primer Pair GC Difference (%)
IGF1_1	48	41	44.5	7
IGF1_2	41	43	42	2
IGF1_3	41	45	43	4
IGF1_4	41	43	42	2
IGF1_5	48	41	44.5	7
IGF1_6	52	40	46	8
IGF1_7	41	45	43	4
IGF1_8	54	48	51	6
IGF1_9	48	52	50	4
IGF1_10	50	41	45.5	9
IGF1_11	45	45	45	0
IGF1_12	46	44	45	2
IGF1_13	41	45	43	4
BP1_1	46	41	43.5	5
BP1_2	48	44	46	4
BP1_3/4	48	48	48	0
BP1_5	52	52	52	0
BP3_1	46	52	49	6
BP3_3	48	48	48	0
BP3_4	44	44	44	0
BP3_6	48	55	51.5	7
BP3_7	45	46	45.5	1
BP3_9	59	52	55.5	7
BP3_10/ 11	42	45	43.5	3
BP3_12	54	40	47	14
BP3_13	48	48	48	0
BP3_14	41	48	44.5	7

Table 5: Putative Amplicon Length, BioMath derived RT-PCR amplicon melting temperature and position of amplicon constituent repeat regions.

Primer Pair Identifier	Amplicon Length (bp)	RT-PCR Amplicon T_M (°C)	Repeat Regions Amplicon Position (bp)
IGF1_1	231	79	-
IGF1_2	243	77	-
IGF1_3	248	81	-
IGF1_4	281	79	-
IGF1_5	175	78	-
IGF1_6	174	80	-
IGF1_7	234	78	-
IGF1_8	72	78	-
IGF1_9	141	85	1-141
IGF1_10	199	81	1-35
IGF1_11	248	80	-
IGF1_12	500	78	-
IGF1_13	250	78	-
BP1_1	138	76	-
BP1_2	205	83	-
BP1_3/4	156	84	-
BP1_5	168	85	-
BP3_1	250	84	-
BP3_3	176	85	-
BP3_4	273	79	-
BP3_6	218	91	-
BP3_7	152	80	-
BP3_9	121	83	-
BP3_10/ 11	168	85	-
BP3_12	205	75	-
BP3_13	154	80	-
BP3_14	269	74	35-269

In Silico PCR

Table 6: Primer pair identifiers and corresponding putative amplicon sequences as derived using UCSC in silico PCR and Ensembl[1] are displayed. Forward primer sequences are detailed in block capital lettering while reverse complement reverse primer sequences are displayed in block capital italics. Probe hybridisation sequences are highlighted in grey and the tagSNP of interest is displayed highlighted in red lettering.

IGF1_1 GGGTCTCTTTCTCTTAGCCTTCTATCTGGcctgaacttctgcatttctctgaatgt cagaataactacacataccacactactaggcttgaagcttagttaagttctctaacgtgatttaaacacagt gcagaaaacacttccatggaagcgtgaacgcttgaaaagactcaaacttagaggatactaattagctact gaga*GAGAAGGCACTATAACAAGACCCAAAGAT*

IGF1_2 ATACCTCAGCATTGGCAATAGATTCTGcaattgatctcagaagagaataattaa gtgtaatagctcacatatgtggtgcttactgtgtgccaggcaatattctaagttaacatattccaatataataa ctaatttaatcctaacaccaactctataaggcaggccccattttacagatgagaaaattgaggaacaaaca ggttaaataacttgct*CAAGATCACTCACCAAGTAAGTTGCAGA*

IGF1_3 GCTAAAGCACATTTGAGATTACACAGACCtgagacctggagaaggtgaga ggttttaatatgaaggctggggaaaaagataaactgccacccctttgggaataatacctattggccaaggc ccatagggatcggcaggttttcttacacatgttttccggttctacttggaacatgagagttatcaccaagtcc ttcatgaaaggtagctagtgttttat*GAGTCGATGGTCAACATCTTGTACTGACT*

IGF1_4 GCTAAAGCTGGAATAATGTGTTAGGTGTGtggcctaggagggtctacattgt gcttttccaaagtgtttttataggatgtaggttttcaggagtgggagagttccctatagagcttggcatttgtt tgtttgttttgctttgttttgaatttggaaagatgatcttaaagaggtatggtaggattaggaaacttttcttcag gcctcagttagtggcctccggtattcaattatactcaggcaatgttaatt*GAGCACTTTCTACAT GCAAATCACTGTC*

IGF1_5 CTCTATATCCCTGGGTGTTACCTGCATAGcatgaagtactatccagaactgac atgcacatgtggaagaacggcagtggggaatgacacattattagatacattggttaccgtgcagaaatgc ttctttatcctgataatatcata*CTCAATAATAGGTCATCACTCAAAGGTGC*

IGF1_6 GCAGACATACCTCTTTCCCTAGAGAGCtctccaggcctggtttcccaggagtg gtggaaataacctggaccttgaatttttttctttttttttttttccacatgactctcaggggactgacacatcaact gaaaacacagttctg*CTTGAAATCCTACTCTGGCTCTTTGTGTTA*

IGF1_7 AGTTGTCCAATATCCTTAAGTGTCTGTGCaaacacttataagtatcatacagtt acttaggaacaatcatttttttccattctaatttttcatcatttaatgcaaaacatcataggcatagaaagatccag ttttcatgcactgtgcatggaaattctggagcaatgtcgtgttgaacatcacaattcaacctggcacgtaac agag*GACACTTCTAGGCTTGGCAATTACCTAGT*

IGF1_8 GCTTTCCCACAGCTAGTGACTGTACCcctaactttgaccagctgtca*CACAG ATGCATAGCAGGCAGTCTAA*

IGF1_9 CTCTCACCTGCCACCATGTAAGATATGtctacttcggccgggcgcggtggctc acgcctgtaatcccagcactttgggaggccgaggcgggtggatcacgaggttagaagatcaagac*CA TCCTGGCTAACACGGTGAAAC*

[1] Ensembl was used to determine the putative amplicon for IGF1_9 primer pair only.

IGF1_10 CGTGTGCCTGTAGTTTCAGCTACTCAggaggctgaggcaggaaaattgcttgaa
cccaggaggcggaggttgcagtgagctgagattgtgccactgcattccagccttggcgacagagtaaa
actctgtctccaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaAAAGATCACTGAG
GACTAATAACAACGGC

IGF1_11 GTTGAGCTAATAGAGAGCTTGAACCTTGGttttcctgagaagggcatgtatag
gtggacaggcccttagtacttttgccaaacctcactcaggcatcttctatgtacccctggtggcgtgaatac
cacagacagctttataatcacacaaagatgagatttgattcatctgcttctatcacagattcattgattgaaag
gagatcctgaaaaatccagcCTTCTGGCCTCTAGATCCTTTCCTAGTAA

IGF1_12 CTGAAGTTCCTCTTGGAAGGCATAACtgggggactttgccttctttcccaaatg
gatggtgttttcagtacccttccccttgtgtcatctttggctccaggcttcccctattgttttgctttcacgtatta
ccgttttggccagactctttcatataacaaactacaaaatagcaccattatactaaaaaacagagttttacat
actgtttgatatatcctgtataattgatatgctaaatttacatagtgctctatatggaaaaaataaaaagagga
aagttactaattaggttgcacattaactcatcatttgaaggaactcttttgagttgaagaaactttctatgtttaa
aacatatgcctaaaaatgattggcctcaaagttgcaactatttgcattattctttttgtaagcatgatgtggaa
aaataaagctttgtgtctaaaataaaatgcatccaacttatatttGGTACAAATGCCACAGAT
GGAATCT

IGF1_13 AGATAATATGGCAGTGCATCTTTCAGCtttcctccttgggggatttttgactgtgg
atagaattaagtgaaggaaataagtcatagacactcttagaattatcacatctaactatgacagaaaacac
gttaagtctgcagaagactgcctataaagttttgttgagagggaataattttaaaaggtacacactgggga
caagaaataaaaagaagtgccatCTTGGGAAGAGGAGTCCAGTATCTTATTC

BP1_1 ATTTCTGCTCTTCCAAAGCTCCTGcgtctgtttttaaagagcatggaaaaatactgc
ctagaaaatgcaaaatgaaataagagagagtagtttttcagctagtttgaagGAGGACGGTTAA
CTTGTATATTCCACCAT

BP1_2 GAGTGCTTTAGGTCTCAGTGAAGTACAGGttctgtagattttattgggagaaac
tgaggactaggccctgcttcacaggcaatgaacagtggggcacacacgagacatgttccctctgggttg
ggctcccctgacatcaggctatgaagcagacagctgtgcacacactgtactgtttaaCACACATG
GGAAGTCATTATTGCAC

BP1_3/4 CACAACTAGAGCTTGAAACCAGAGCACgtagttggggaaggagcttgggtca
cccagtggagcccgctcattgcacggtcttggcaggacgtgctctgggagaagaaggaagatgttcca
gggcacaCATAGCTTAGTGGAGACTCAAGGAGGAAG

BP1_5 TCCTGGAGACTCTAGCTCCCTATCTTGggtccccctttaaccctccaaccccat
agctgccaggctgccatcctctctgccatgaccctagggctggtgtgtaaactcagcactcagactcatc
aagaggagcagccagaccGATGTGGGACAGGGACAGACTGATA

BP3_1 AACACGCTTATAAGAGCTTGGTGTCCagctcagatgggaaaactgaattattac
ctgcaaagcgttttgcctcaggggctcagaatcatgcaagcatgttggtggcttgttttctaggtccccgtt
acatctctaaaactcaaggtctacacagaccctgtgcatcttgctggtgcctgcctaagtgagtccttttcac
cccgggtgaacacagcagcacatggatgCAGACAATGCCAAAGTCCCACTC

BP3_3 GACTCTGCTATGCTGAGAAAGCACAACagaaatttcagctaaggcaacacaa
gagccatgcgtgcctaggcccgctgagtgtgcgcctgtgcatgcgtgtgggtgtaacttcctgctccaag
aaagcgggggtgggggcagttttgt*GTTTCTAGGAGTATAAGGTGGACACCAGC*

BP3_4 TACACCGCAAGTCTCCAATTAAGAGTGgaccggcaagcgaatgcgtccttaa
ggcagggcttttcaaatattttaaatgacaactgttcttcttgtcttgggtattctccctgattatgttttcaaatg
gagtttacacccatgacaaaagaataataaagacaataaactgggcactgctgaaacgtaattaaccaaa
tagtcctttaataggcaattttcattgtcatttttgaaggcgacttt*CGTATTACTGTATTCTCGA*
TGCGTTGG

BP3_6 TGTCGTCTACAAGAACCAAGGTGTGcccggccaccccggcactccaggccac
ctcagcacccccggtcaccccagtcactcctggccaactcagcacccccgatctccttgaccccgccgc
ccttaccccttccgctctcggggtgaggtctccctgcggcgagccggtgtcggggaaactggcatacag
cgctcc*GCATTCGTGTGTACCTCGTGGA*

BP3_7 CATCATACTACTCACTACATGGTGGTTGCtctacctcaagaagttatctgtttga
aagtcaatctgacaggtgctcttcctaagtagctgcaactaaagaaggcagacaaacgcttcagtgccc
CTCAGTAGGTGAAGGCAGTATGCTTCAT

BP3_9 GCCCTGAGATATCCAGCACAGCctgcaggctaatggcactaggcctgcaagtgca
gggctggggtctcaactcatgttttcaaacaaagcaacgagta*CCAGATGCTGCTCACAGT*
GTTCA

BP3_10/
11 ATTACTTGTGATGCCTCTGAATGTGGaggctgactctccctgtctctctgtccctc
ctaccccacggggccgcagcaaaagccatcctgggccttcgactgggccatgtcttcaggaagattcct
gaagaggagggccc*GAAATACCTGCCTTTATAGGTTCCCAGAG*

BP3_12 GTGAGCTCCTTTCCTCAGTCATGGccacagttgtatcatatagcatctctaacatttc
atctaggattatctagtatagatcttactatatttggggctatgttgtatacaatgttaacaagaacatatcttct
ctgcatatatgtgtgaattataaagaaaagcatgagaatga*CTCTAAGTTCAACAAACATG*
GGTGAATCTC

BP3_13 ACACACCACAATACCAGTCCTCTGAACacttacagaaccggcttgctgctcatc
acatacaacacgtgataagggtatcattttagatgtttcatcaacaaatcagacatggcgctagacaggag
GTAACTCGACAGGAGGTAACTCTTCGACT

BP3_14 GTTGATCATAGGTATTGTGTCAGGGTTTCccagagaaacagaactaacaggt
tatatctatctgtctgtctatctatctatctatctatctatctatctatctatctatctatcatctatgtatctatc
tctgtctaatctacctatctatctaatctatcaatctatcatctatctagtctatctacctacttatctctaatcacc
caccaaaaaagacttattacaaaaaattg*GCTAAGTATGGAGGCTAAGTCTCACCA*
TC

Table 7: IGF Multiplex Probe Set Data: Gene name, Tagger Bin number, dbSNP ref sequence identifier, SNP allele and given Probe Identifier are displayed

Gene Name	Tagger Bin Number	Reference Sequence tagSNP Identifier	SNP Allele	Probe Identifier
IGF1	1	rs972936	T	IGF1_1T
			C	IGF1_1C
	2	rs17882461	A	IGF1_2A
			T	IGF1_2T
	3	rs11111262	G	IGF1_3G
			A	IGF1_3A
	4	rs2033178	A	IGF1_4A
			G	IGF1_4G
	5	rs2946834	A	IGF1_5A
			G	IGF1_5G
	6	rs35767	A	IGF1_6A
			G	IGF1_6G
	7	rs12821878	G	IGF1_7G
			A	IGF1_7A
	8	rs1019731	C	IGF1_8C
			A	IGF1_8A
	9	rs17884646	C	IGF1_9C
			T	IGF1_9T
	10	rs12316064	C	IGF1_10C
			T	IGF1_10T
	11	rs1520220	G	IGF1_11G
			C	IGF1_11C
	12	rs3730204	A	IGF1_12A
			G	IGF1_12G
	13	rs6214	C	IGF1_13C
			T	IGF1_13T

IGFBP1	1	rs7454	C	BP1_1C
			G	BP1_1G
	2	rs1874479	A	BP1_2A
			G	BP1_2G
	3	rs3828998	T	BP1_3T
			C	BP1_3C
	4	rs9658194	C	BP1_4C
			A	BP1_4A
	5	rs9658238	A	BP1_5A
			G	BP1_5G
IGFBP3	1	rs3793345	T	BP3_1T
			C	BP3_1C
	3	rs10255707	T	BP3_3T
			C	BP3_3C
	4	rs2132571	T	BP3_4T
			C	BP3_4C
	6	rs11765572	G	BP3_6G
			A	BP3_6A
	7	rs12540724	A	BP3_7A
			G	BP3_7G
	9	rs2453839	T	BP3_9T
			C	BP3_9C
	10	rs35751739	C	BP3_10C
			T	BP3_10T
	11	rs35496550	T	BP3_11T
			-	BP3_11-
	12	rs6670	T	BP3_12T
			A	BP3_12A
	13	rs13223993	G	BP3_13G
			A	BP3_13A
	14	rs10282088	C	BP3_14C
			A	BP3_14A

Table 8: Panjokovich consensus predicted T_M (°C) for IGF Multiplex probe set. The predicted probe annealing temperature for each probe pair, mean primer pair T_A and intra-primer pair T_A difference as predicted using Primer3 predictions are displayed.

Probe Identifier	Probe T_A (°C)	Mean Probe Pair T_A (°C)	Intra-probe pair T_A Difference (°C)
IGF1_1T	59.51	59.75	0.49
IGF1_1C	60.00		
IGF1_2A	60.38	59.97	0.82
IGF1_2T	59.56		
IGF1_3G	59.90	60.11	0.43
IGF1_3A	60.33		
IGF1_4A	59.27	60.20	1.88
IGF1_4G	61.15		
IGF1_5A	60.00	60.08	0.17
IGF1_5G	60.17		
IGF1_6A	60.10	60.01	0.17
IGF1_6G	59.93		
IGF1_7G	60.23	60.03	0.4
IGF1_7A	59.83		
IGF1_8C	60.77	61.11	0.68
IGF1_8A	61.45		
IGF1_9C	59.84	60.17	0.66
IGF1_9T	60.50		
IGF1_10C	59.70	60.22	1.05
IGF1_10T	60.75		
IGF1_11G	59.81	59.94	0.27
IGF1_11C	60.08		
IGF1_12A	59.90	60.14	0.49
IGF1_12G	60.39		
IGF1_13C	59.53	60.11	1.16
IGF1_13T	60.69		

BP1_1C	60.63	60.14	0.98
BP1_1G	59.65		
BP1_2A	59.45	60.21	1.52
BP1_2G	60.97		
BP1_3T	60.57	59.26	2.61
BP1_3C	57.96		
BP1_4C	60.84	60.54	0.59
BP1_4A	60.25		
BP1_5A	59.79	60.23	0.89
BP1_5G	60.68		
BP3_1T	59.50	59.96	0.92
BP3_1C	60.42		
BP3_3T	60.53	59.4	2.26
BP3_3C	58.27		
BP3_4G	60.51	60.19	0.63
BP3_4A	59.88		
BP3_6T	60.13	60.46	0.66
BP3_6C	60.79		
BP3_7C	60.08	60.00	0.15
BP3_7T	59.93		
BP3_9T	59.61	59.84	0.46
BP3_9A	60.07		
BP3_10G	58.81	59.56	1.51
BP3_10A	60.32		
BP3_11C	59.46	59.93	0.94
BP3_11A	60.40		
BP3_12T	59.45	59.80	0.7
BP3_12C	60.15		
BP3_13T	60.00	60.29	0.59
BP3_13C	60.59		
BP3_14T	61.32	61.16	0.32
BP3_14C	61.00		

Table 9: % GC for all probes, mean probe pair GC% and Intra-probe pair difference in the final IGF Multiplex set as determined using Primer3 are displayed.

Probe Identifier	Probe (GC %)	Mean Probe Pair GC (%)	Intra-probe pair GC Difference (%)
IGF1_1T	39.13	40.02	1.78
IGF1_1C	40.91		
IGF1_2A	32.00	32.66	1.33
IGF1_2T	33.33		
IGF1_3G	52.63	50.12	5.01
IGF1_3A	47.62		
IGF1_4A	40.91	45.45	9.09
IGF1_4G	50.00		
IGF1_5A	30.00	31.66	3.33
IGF1_5G	33.33		
IGF1_6A	47.62	50.12	5.01
IGF1_6G	52.63		
IGF1_7G	36.00	34.66	2.67
IGF1_7A	33.33		
IGF1_8C	43.48	42.57	1.81
IGF1_8A	41.67		
IGF1_9C	55.00	52.50	5.00
IGF1_9T	50.00		
IGF1_10C	68.75	59.37	10.86
IGF1_10T	57.89		
IGF1_11G	52.63	51.31	2.63
IGF1_11C	50.00		
IGF1_12A	37.50	37.50	0
IGF1_12G	37.50		
IGF1_13C	37.50	37.98	0.96
IGF1_13T	38.46		

BP1_1C	39.13	39.13	0
BP1_1G	39.13		
BP1_2A	50.00	51.47	2.94
BP1_2G	52.94		
BP1_3T	66.67	70.00	6.66
BP1_3C	73.33		
BP1_4C	55.56	54.09	2.93
BP1_4A	52.63		
BP1_5A	73.33	71.96	2.74
BP1_5G	70.59		
BP3_1T	52.63	48.81	7.63
BP3_1C	45.00		
BP3_3T	55.56	64.44	17.77
BP3_3C	73.33		
BP3_4G	39.13	43.37	8.49
BP3_4A	47.62		
BP3_6T	78.57	71.64	13.86
BP3_6C	64.71		
BP3_7C	45.45	52.13	13.37
BP3_7T	58.82		
BP3_9T	32.00	37.74	11.48
BP3_9A	43.48		
BP3_10G	70.59	68.63	3.92
BP3_10A	66.67		
BP3_11C	59.09	60.49	2.81
BP3_11A	61.90		
BP3_12T	33.33	31.66	3.33
BP3_12C	30.00		
BP3_13T	52.63	52.50	0.25
BP3_13C	52.38		
BP3_14T	34.30	32.00	4.60
BP3_14C	29.70		

Table 10: IGF Multiplex Probe SNP position. Probe identifier, probe length, SNP position closest to the sequence terminus and position designation of central or skewed.

Probe Identifier	Probe length (bp)	SNP position (closest to probe terminus) (bp)	SNP Position Designation* (central / skewed)
IGF1_1T	23	12	central
IGF1_1C	22	9	central
IGF1_2A	25	3	Skewed
IGF1_2T	24	2	skewed
IGF1_3G	19	9	central
IGF1_3A	21	10	central
IGF1_4A	22	11	central
IGF1_4G	20	10	central
IGF1_5A	30	15	central
IGF1_5G	27	13	central
IGF1_6A	21	7	central
IGF1_6G	19	6	central
IGF1_7G	25	11	central
IGF1_7A	27	13	central
IGF1_8C	23	4	Skewed
IGF1_8A	24	5	skewed
IGF1_9C	20	2	Skewed
IGF1_9T	22	3	skewed
IGF1_10C	16	8	central
IGF1_10T	19	8	central
IGF1_11G	19	3	Skewed
IGF1_11C	20	4	skewed
IGF1_12A	24	12	central
IGF1_12G	24	13	central
IGF1_13C	24	7	central
IGF1_13T	26	10	central

BP1_1C	23	6	central
BP1_1G	23	6	central
BP1_2A	20	10	central
BP1_2G	17	9	central
BP1_3T	18	8	central
BP1_3C	15	5	central
BP1_4C	18	5	central
BP1_4A	19	6	central
BP1_5A	15	4	central
BP1_5G	17	8	central
BP3_1T	19	7	central
BP3_1C	20	10	central
BP3_3T	18	9	central
BP3_3C	15	4	central
BP3_4G	23	7	central
BP3_4A	21	7	central
BP3_6T	14	4	central
BP3_6C	17	6	central
BP3_7C	22	11	central
BP3_7T	17	7	central
BP3_9T	25	12	central
BP3_9A	23	11	central
BP3_10G	17	5	central
BP3_10A	18	5	central
BP3_11C	22	10	central
BP3_11A	21	10 (del)	central
BP3_12T	30	14	central
BP3_12C	30	13	central
BP3_13T	19	8	central
BP3_13C	21	11	central
BP3_14T	35	15	central
BP3_14C	37	15	central

* SNPs in the interquartile range were designated as “central”, while those in the 25th or 75th percentile were designated as being in a “skewed” position.

1.2 Chapter 4: Multiplex PCR Amplification

1.2.1 Cell Culture

OE21 human Caucasian oesophageal squamous cell carcinoma derived cell line was removed from nitrogen storage, thawed, inoculated into RPMI 1640 media with L-glutamine and NANCO2. This medium was supplemented and incubated at 37°C in a 5% CO2 enriched atmosphere. Cell cultures were monitored, fed and split as follows:

Table 3 - Media Enrichment for Culture of cell lines

	Ingredients	Volume	Function
1	10% (v/v) Foetal calf serum (FCS)	50 ml	Nutritional
2	Penicillin (5units/ml)/streptomycin (0.005mg/ml)	500 µl	Antibacterial
3	Amphotericin B (2.50 units/ml)	500 µl	Antifungal
4	Glutamine	1 ml	Nutritional

Note good cell culture practices were followed throughout these procedures to minimise contamination.

Media inoculation

1. Media containing 10% FCS was warmed to 37^{O}C and a water bath to 70^{O}C.
2. A stock vial of OE21 was removed from nitrogen storage and swirled gently in the water-bath until the contents had thawed.
3. The contents of the vial was extracted and placed in a 15ml centrifuge tube.
4. 1 drop of media was added using a plastic Pasteur pipett and the mixture swirled. This was done over the course of 15 minutes until a volume of

approximately 5ml was reached. A further 5mls was then added over the course of 1 minute.

5. The mixture was centrifuged at 4^{O}C, 100g (1300rpm) for 5 mins.
6. Supernatant was decanted and the pellet was resuspended in 1ml of media.
7. The contents of the centrifuge tube were added to a T25 flask and a further 6mls of media added.
8. The flask was swirled gently to ensure an even distribution of cells and placed in a 37^{O}C incubator.
9. The contents of the T25 flasks were checked the next day to ensure that cells have grown in a thin film across the bottom of the flask.

Feeding cell cultures

Once established – cultures were checked daily and the media replaced every two days. Media should appear light pink in colour, turning slightly yellow when cells require feeding. Media should always be clear as turbidity is an indication of infection, contamination or non-adherence of cells.

1. Stock media was warmed to 37^{O}C, and a waste beaker prepared in accordance with disinfection protocols
2. T25 flasks (maximum of two at a time) were removed from the incubator and media decanted into the waste beaker.
3. 5ml of warmed media (37^{O}C) was added to each flask, these were gently swirled before reincubation.

Splitting cell cultures

Cell cultures were checked daily, to ensure they were growing adequately and adhering to the flask surface. Percentage coverage was also assessed and split once a confluence state of 75-80% coverage had been reached. Cells were split using the trypsin/EDTA detachment method as follows:

1. Trypsin/EDTA (10x) was thawed and warm sterile PBS and media warmed to 37^{O}C. 1x trypsin/EDTA was prepared by adding 1ml of trypsin concentrate to 9mls of PBS.
2. Media was decanted from flasks and residual media rinsed using sterile PBS.
3. 1x trypsin solution (5ml for a T25 flask) was added and flasks placed in a 37^{O}C incubator for 1 minute.
4. Flasks were removed and tapped vigoursly and checked using a microscope to determine if the cells were becoming detached.
5. 1 minute incubation followed by tapping and checking was repeated until all the cells were in suspension.
6. Once cells had become detached the suspension was removed and added to the centrifuge tube containing 5ml of media (which acts to neutralise the trypsin) and centrifuged at 4^{O}C, 100g (1300rpm) for 5 mins.
7. Supernatant was decanted and pelletts resuspended in 5mls of media.
8. Each T25 was split into two T75's and 1ml of cell suspension added followed by 9ml of media.
9. The solution was gently swirled to ensure adequate dispersion of cells and flasks incubated as before.

1.2.2 DNA Quantification and Qualification

All singleplex amplifications were carried out using OE21 cell line template.

DNA was extracted using the Qiagen QIAamp DNA Mini Kit according to the manufacturers instructions. Resultant DNA yield and purity was tested using the spectrophotometric molar extinction coefficients absorption method first described by Warburg and Christian using the following equation:

- Quantitation (µg/ml): (A260)(dilution factor)(50 ngDNA/µl)

Where 1 spectrophotomic unit equals 50 ng/µl of double stranded DNA at OD = 260nm

Nucleic acid degradation and PCR inhibition can result from insufficient purification of DNA. Deposition of cellular contaminants including proteins, polysaccharides and salts (DNase, RNase, heparin, bile salts) can have debilitating effects on downstream applications, it is therefore essential to test for potential contamination and perform further purification if required (Holodniy et al. 676-79). Protein contamination is a good indication of overall sample purity and was used to assess purity of the sample.

Protein peptide bonds absorb maximally at A228, while polypeptides containing multiple aromatic residues absorb maximally at A280. Measurement at both wavelengths provides a more broad ranging assessment of sample purity.

- Purity (score): A260/A228

  A260/A280

Samples displaying scores of >2 for A260/A228 and between 1.7 and 1.9 for A260/A280 were used for subsequent experiments.

All spectrophotometric measurements were carried out using Eppendorf BioPhotometer (Eppendorf UK Limited, Cambridge, UK) and Hellma Quartz 10mm (Hellma UK LTD, Essex, UK) cuvettes. to ensure accurate spectrophotomic readings; all used dilutions provided OD readings between 0.1 and 1.0. Ten readings were performed for both absorbance wavelengths (A260 and A280) and all were carried out in duplicate.

1.2.3 Agarose Gel electrophoresis

The results of the PCR reaction were prepared, separated by size using agarose gel electrophoresis Sub-Cell® GT kit (Bio-Rad Laboratories, Inc., Hertfordshire, UK) and visualised using ethidium bromide /UV illumination Syngene Gene Genius Bioimaging System (Synoptics Ltd, Cambridge, UK)

1. 1.2µl of X6 Blue / Orange loading dye was gently mixed with 2µl of 100bp ladder for size standard preparation and 5µl aliquot from each completed PCR reaction tube for sample preparation.

2. 2% (w/v) agarose gel was made as follows: 1g of agarose powder was added to 50ml of 1xTAE buffer (see Appendix H for TAE preparation protocol).
3. This was heated till it boiled and allow cool to a temperature of just below 70^{o}C
4. The molten gel was poured into the gel tray and the comb placed into position. This was allowed to set for 20 minutes on a flat surface.
5. The comb was removed and the mould (containing the gel) was placed into the electrophoresis tank. Note – Care was taken to position the gel so that sample

wells were closest to the negative charge as DNA is itself negatively charged and therefore will run to the positive through the gel.

6. 1XTAE buffer was poured into the tank until the gel was fully covered by buffer.
7. Pre-prepared samples and size standards were loaded into their respective wells.
8. The tank was connected to a power source and ran at 100 Volts for 45 minutes for the small 50ml gels (or for 2 hrs 30 mins for larger 100ml gels).
9. Each gel was placed in 0.5 µg/ml of ethidium bromide solution for 30 minutes.
10. UV illumination of processed / stained gels allowed size standard and products to be visualised.
11. Bands were sized by comparison to 100bp size standard (100–1,500bp)

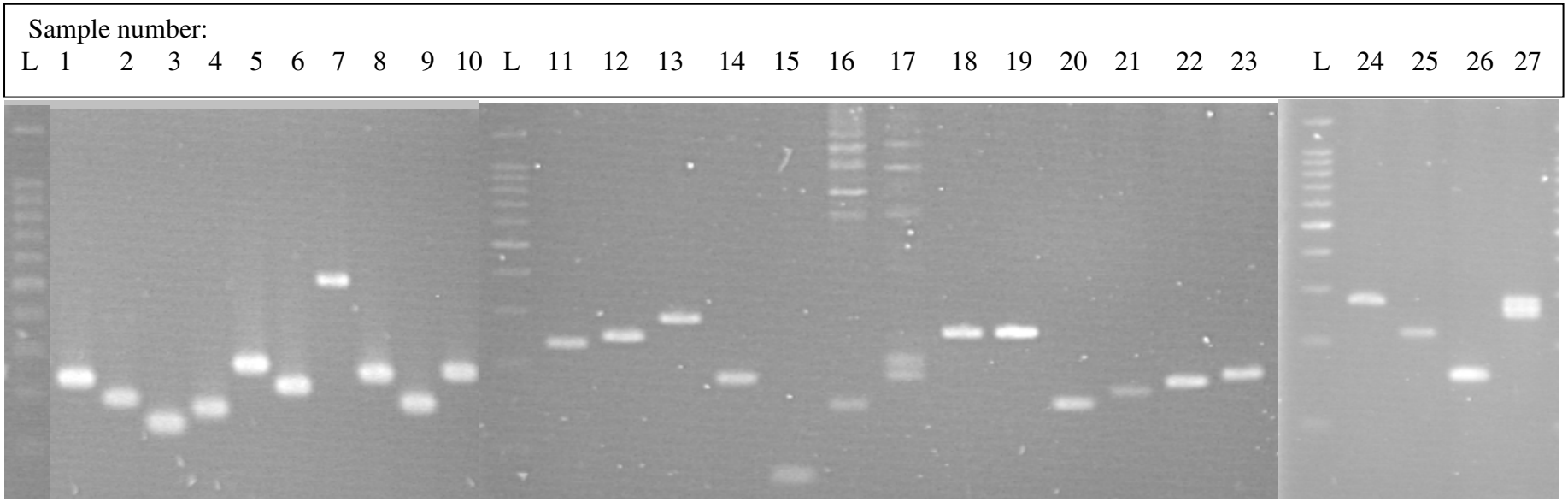


Figure 1: Singleplex Primer Pair Amplification using the Qiagen Multiplex PCR kit. All pairs were amplified singly using uniform reaction conditions as specified. Target specific amplification may be observed for all pairs bar IGF1_9, IGF1_10 and BP3_14 sample numbers 16, 17 and 27 respectively. All primer pair Identifiers and corresponding sample numbers are detailed in table 11 and lanes containing ladder labelled "L".

Table 11: Singleplex Primer Pair Amplification using the Qiagen Multiplex PCR kit. Sample numbers and corresponding primer pair identifiers relating to figure 1 are displayed.

Sample number	Primer Pair Identifier
1	BP3_12
2	BP3_10/11
3	BP3_9
4	BP3_7
5	BP3_1
6	BP1_2
7	IGF1_12
8	IGF1_7
9	IGF1_5
10	IGF1_3
11	IGF1_1
12	IGF1_2
13	IGF1_4
14	IGF1_6
15	IGF1_8
16	IGF1_9
17	IGF1_10
18	IGF1_11
19	IGF1_13
20	BP1_1
21	BP1_3/4
22	BP1_5
23	BP3_3
24	BP3_4
25	BP3_6
26	BP3_13
27	BP3_14

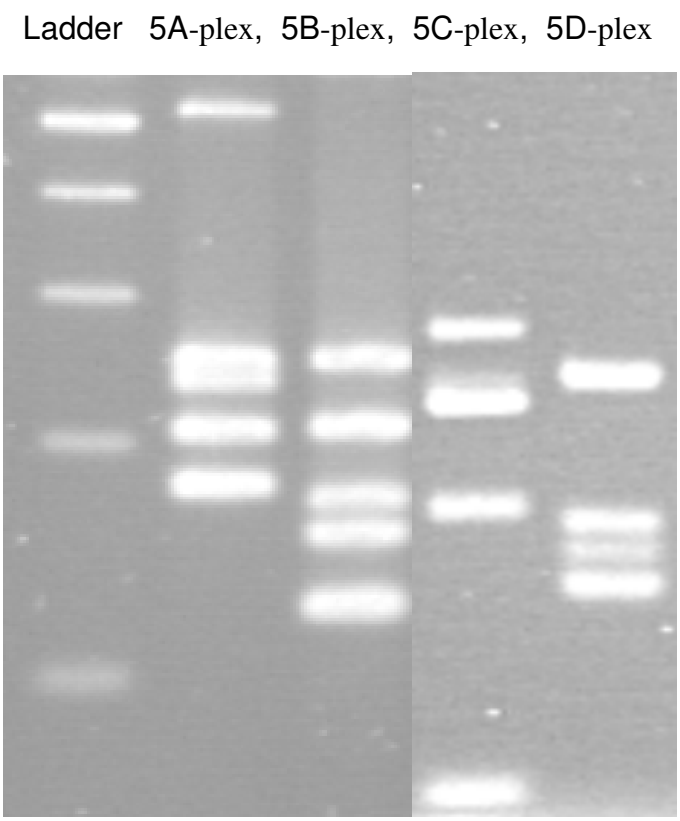


Figure 2: Multiplex PCR amplification of four 5-plexes using MetaPhor agarose gel electrophoresis. Multiplex PCR amplification appears to have been relatively successful with clear resolution of amplicons ≥16bp (5B-plex BP3_7 and BP3_10/11; 152 and 168bp respectively).

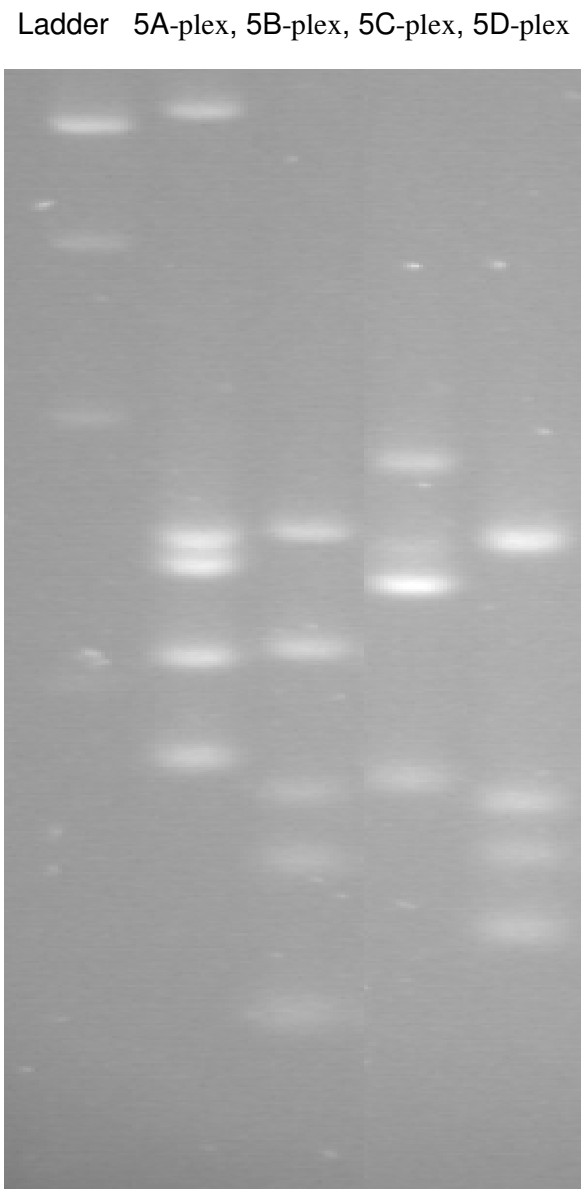


Figure 3: Separation of four 5-multiplex PCR products using higher resolution MetaPhor gel protocol. Use of longer separation times with increased MetaPhor gel and buffer concentrations resulted in increased resolution; allowing 12bp resolution of IGF1_3 and IGF1_7 amplicon bands (243 and 231bp respectively, 5A-plex). However band sharpness and sensitivity was significantly diminished especially for lower molecular weight amplicons.

Table 12: 5-plex primer pair and target description. Multiplex Identifier, constituent primer pair identifier and amplicon sizes for 5A-plex, 5B-plex, 5C-plex and 5D-plex, corresponding to MetaPhor gels figures 3 and 4 are detailed.

Multiplex Identifier	Primer Pair ID	Amplicon size (bp)
5A-plex	IGF1_12	500
	IGF1_3	248
	IGF1_7	234
	BP1_2	205
	IGF1_5	155
5B-plex	BP3_1	250
	BP3_12	205
	BP3_10/11	168
	BP3_7	152
	IGF1_8	72
5C-plex	IGF1_4	281
	IGF1_2	243
	IGF1_1	231
	IGF1_6	174
	BP3_9	121
5D-plex	IGF1_11	248
	IGF1_13	250
	BP1_1	138
	BP1_3/4	156
	BP1_5	168

Table 13: 14-plex amplicon size determination; *in silico* predicted amplicon size and Experion derived size approximations are displayed along with deviation (bp) between the two.

No.	Primer Pair Identifier	Predicted Amplicon size (bp)	Experimental amplicon size approximation (bp)	Deviation between predicted and experimental amplicon size approximation (bp)
1	IGF1_8	72	84	6
2	BP3_9	121	128	7
3	BP1_1	138	141	3
4	BP3_7	152	158	6
5	BP1_3/4	156	165	10
6	BP3_10/11	168	175	7
7	BP3_3	176	188	12
8	BP1_2	205	215	10
9	BP3_6	218	219	1
10	IGF1_1	231	236	5
11	IGF1_13	250	249	1
12	BP3_4	273	270	3
13	IGF1_4	281	278	3
14	IGF1_12	500	500	0

Table 14: 6-plex amplicon size determination; *in silico* predicted amplicon size and Experion derived size approximations are displayed along with deviation (bp) between the two.

No.	Primer Pair Identifier	Predicted Amplicon size (bp)	Experimental amplicon size approximation (bp)	Deviation between predicted and experimental amplicon size approximation (bp)
1	BP3_13	154	164	10
2	BP1_5	168	177	9
3	IGF1_5	175	188	13
4	BP3_12	205	216	11
5	IGF1_7	234	243	9
6	IGF1_3	248	257	9

Table 15: 5-plx amplicon size determination; *in silico* predicted amplicon size and Experion derived size approximations are displayed along with deviation (bp) between the two.

No.	Primer Pair Identifier	Predicted Amplicon size (bp)	Experimental amplicon size approximation (bp)	Deviation between predicted and experimental amplicon size approximation (bp)
1	BP1_3/4	156	166	10
2	BP3_10/11	168	177	9
3	IGF1_6	174	191	17
4	IGF1_2	243	253	10
5	BP3_1	250	261	11

Table 16: 2-plex amplicon size determination; *in silico* predicted amplicon size and Experion derived size approximations are displayed along with deviation (bp) between the two.

No.	Primer Pair Identifier	Predicted Amplicon size (bp)	Experimental amplicon size approximation (bp)	Deviation between predicted and experimental amplicon size approximation (bp)
1	IGF1_9	141	147	6
2	IGF1_11	248	260	12

Table 17: 14-plex primer pair yield (ng/µl)

No.	Primer Pair Identifier	Amplicon Concentration (ng/µl)
1	IGF1_8	8.24
2	BP3_9	6.58
3	BP1_1	6.16
4	BP3_7	5.04
5	BP1_3/4	6.54
6	BP3_10/11	4.62
7	BP3_3	5.1
8	BP1_2	7.46
9	BP3_6	6.4
10	IGF1_1	6.5
11	IGF1_13	4.36
12	BP3_4	5.28
13	IGF1_4	6.1
14	IGF1_12	6.64

Table 18: 6-plex primer pair yield (ng/µl)

No.	Primer Pair Identifier	Mean Amplicon Concentration (ng/µl)
1	BP3_13	9.13
2	BP1_5	12.08
3	IGF1_5	11.08
4	BP3_12	12.45
5	IGF1_7	7.26
6	IGF1_3	8.42

Table 19: 5-plex primer pair yield (ng/µl)

No.	Primer Pair Identifier	Mean Amplicon Concentration (ng/µl)
1	BP1_3/4	11.31
2	BP3_10/11	11.46
3	IGF1_6	8.99
4	IGF1_2	8.82
5	BP3_1	12.31

Table 20: 2 -plex primer pair yield (ng/µl)

No.	Primer Pair Identifier	Mean Amplicon Concentration (ng/µl)
1	IGF1_9	12.4
2	IGF1_11	10.48

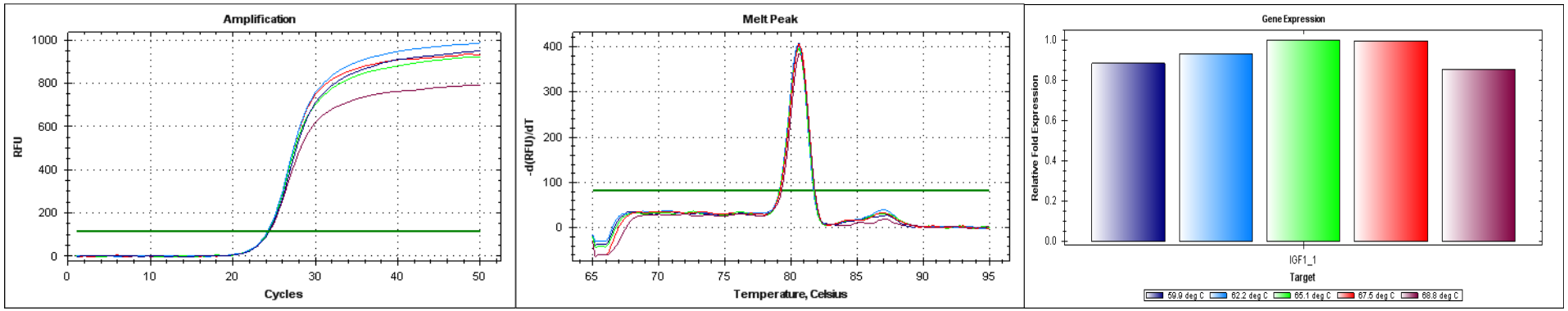


Figure 4: IGF1_1 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

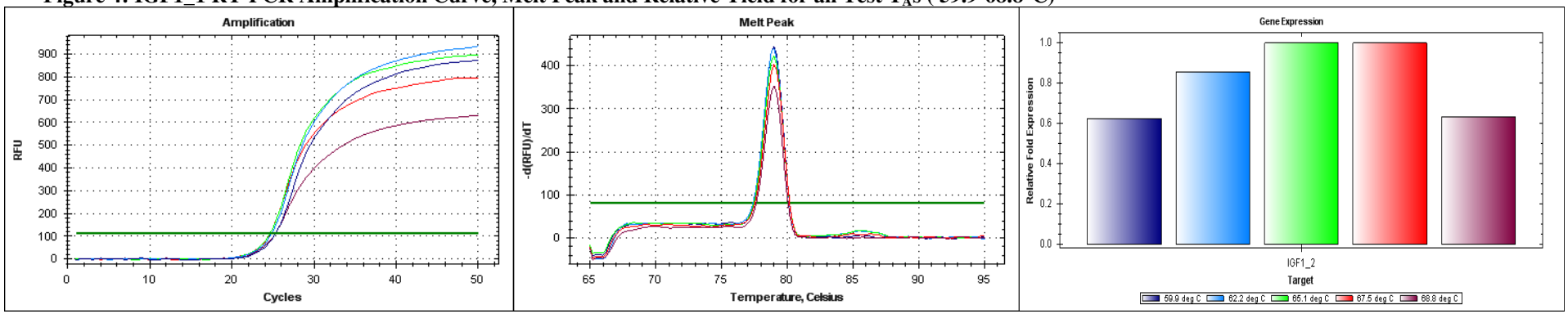


Figure 5: IGF1_2 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

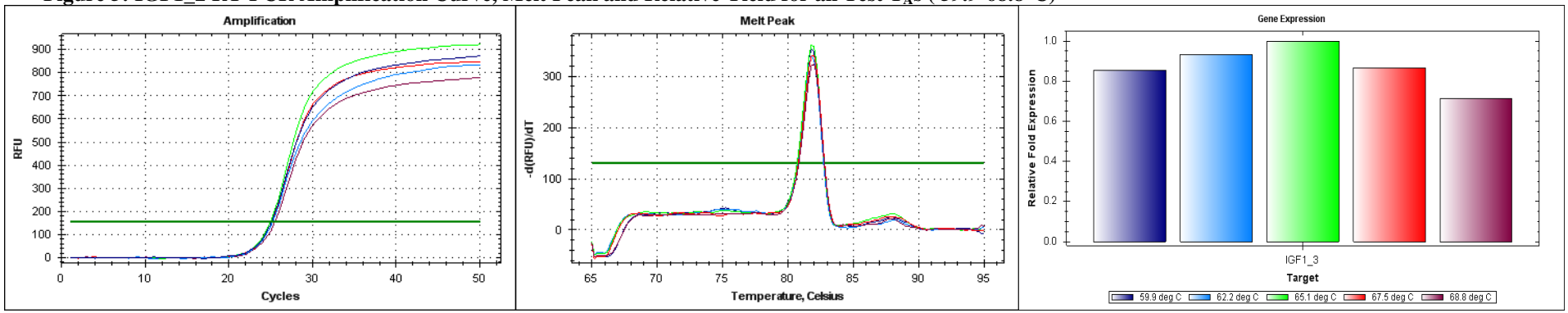


Figure 6: IGF1_3 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

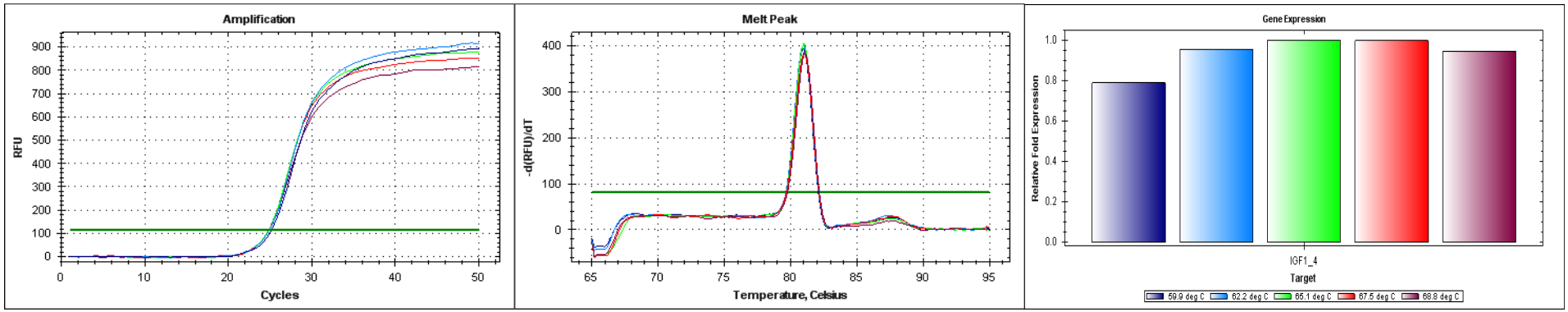


Figure 7: IGF1_4 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

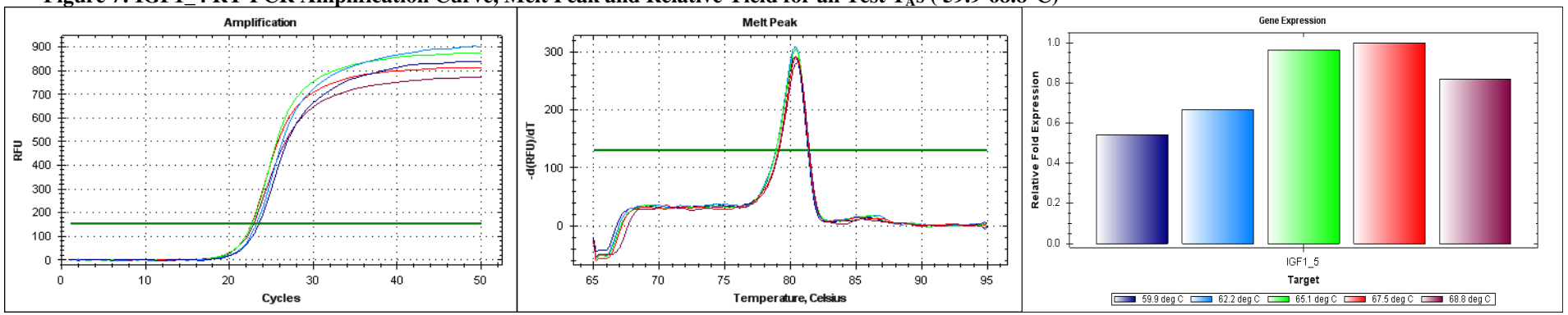


Figure 8:IGF1_5 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

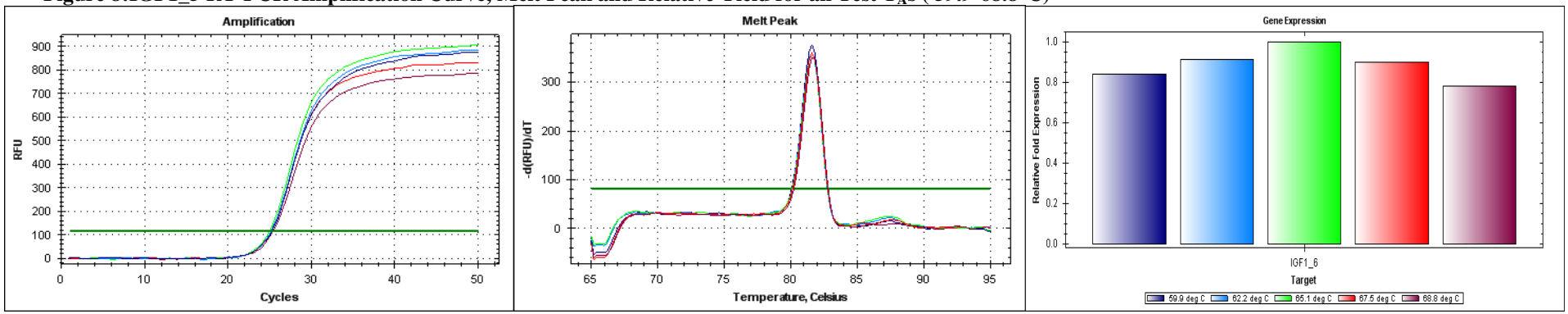


Figure 9: IGF1_6 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

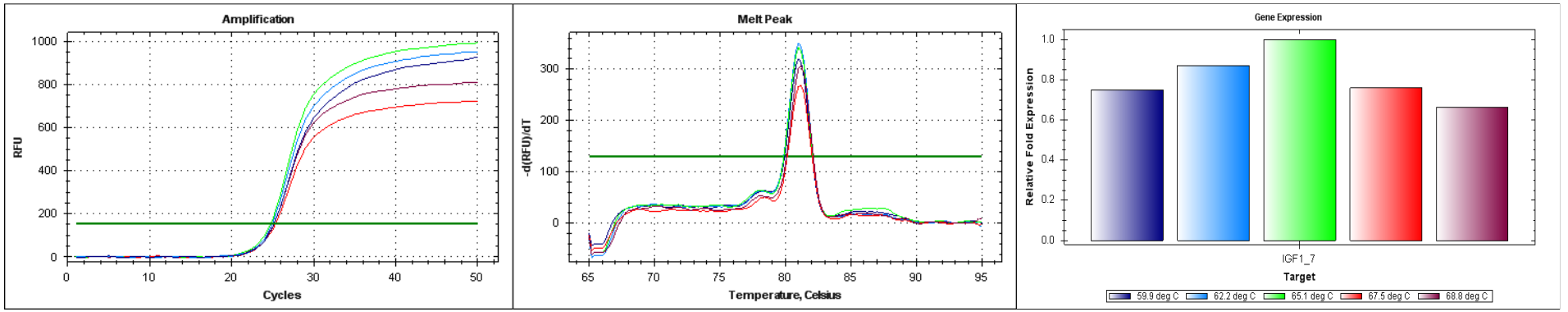


Figure 10: IGF1_7 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

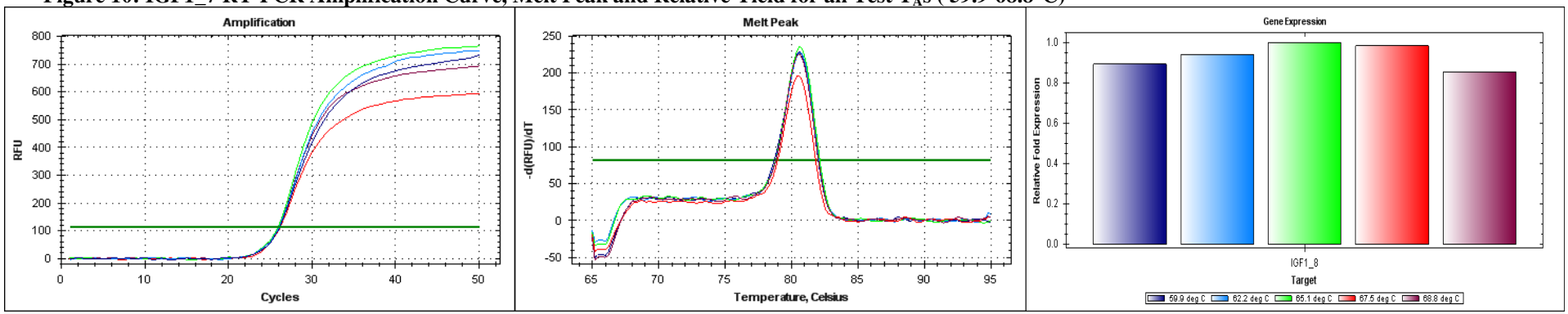


Figure 11: IGF1_8 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

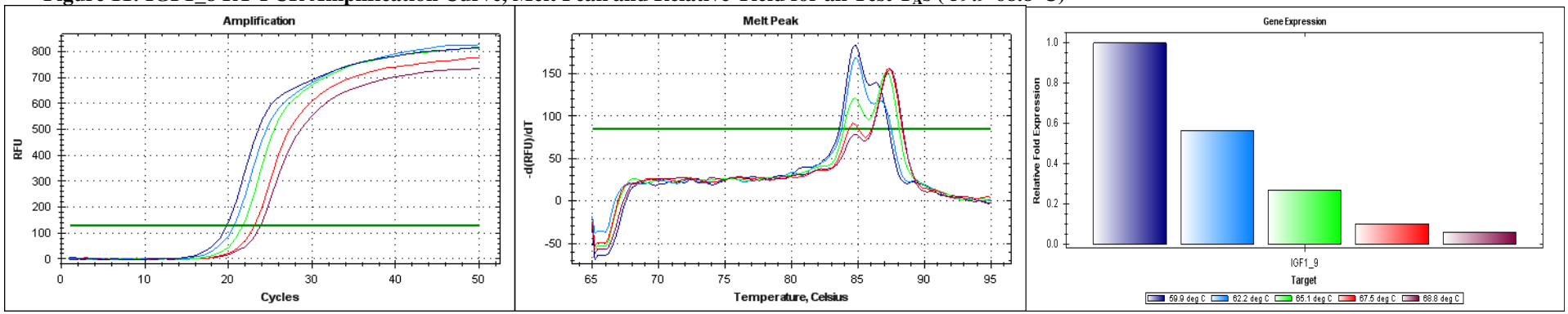


Figure 12: IGF1_9 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

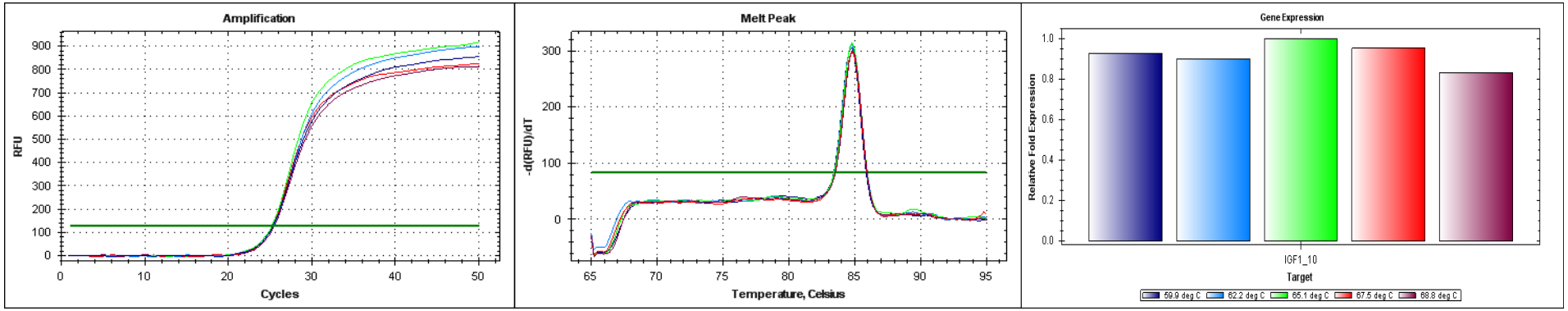


Figure 13: IGF1_10 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

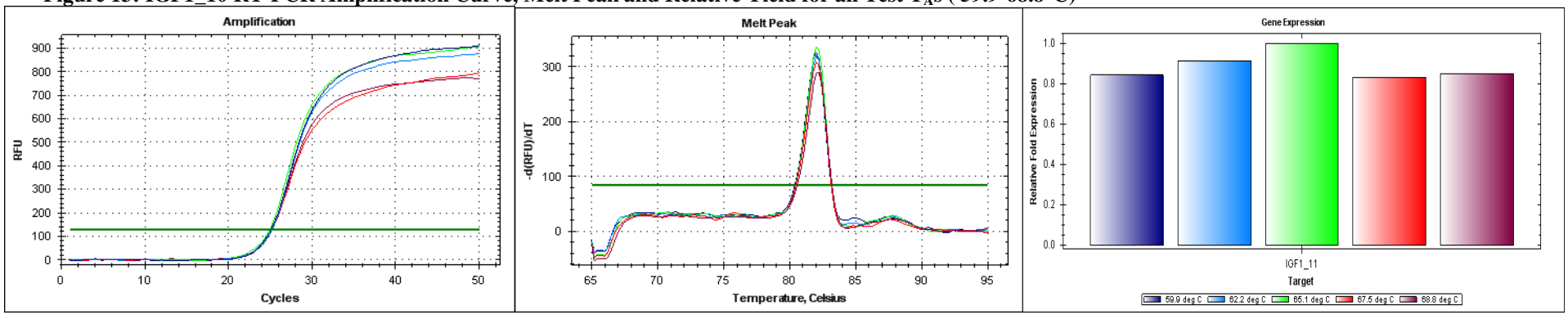


Figure 14: IGF1_11 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

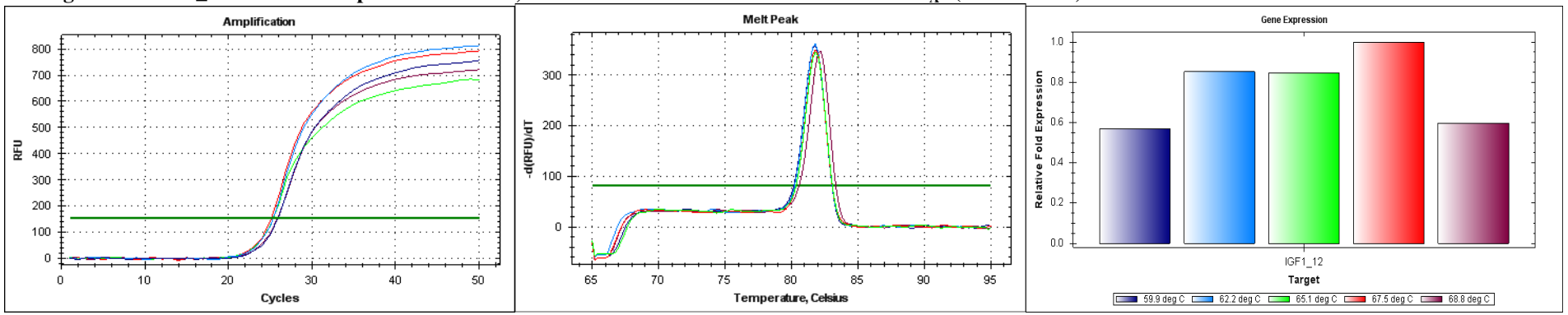


Figure 15: IGF1_12 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

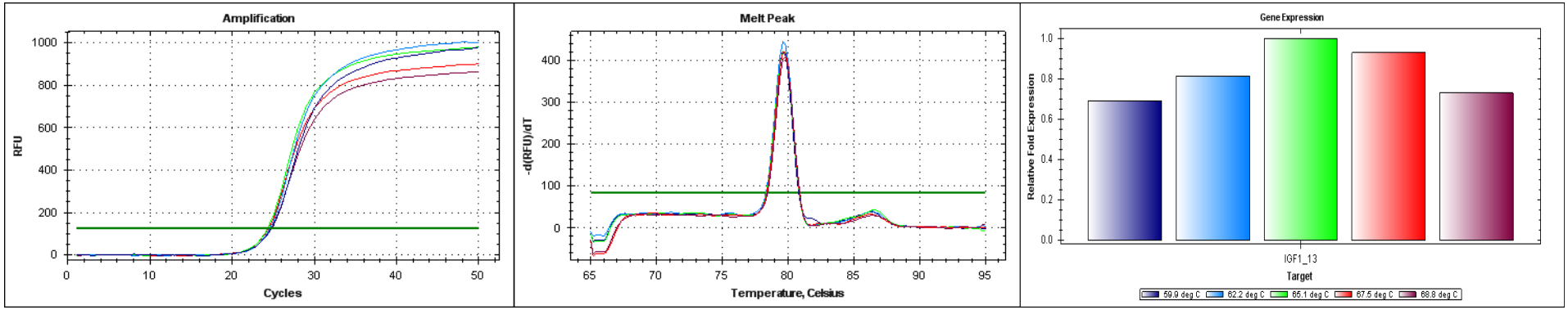


Figure 16: IGF1_13 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

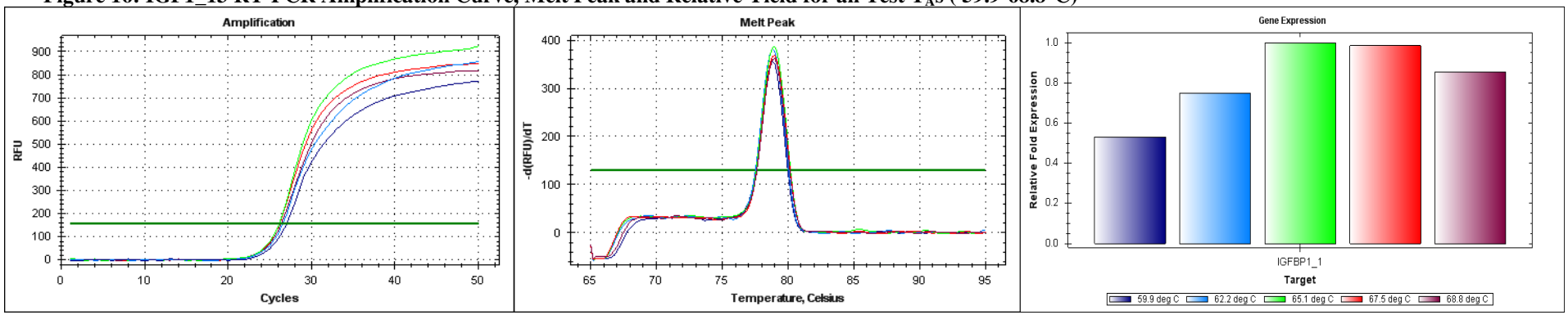


Figure 17: BP1_1 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

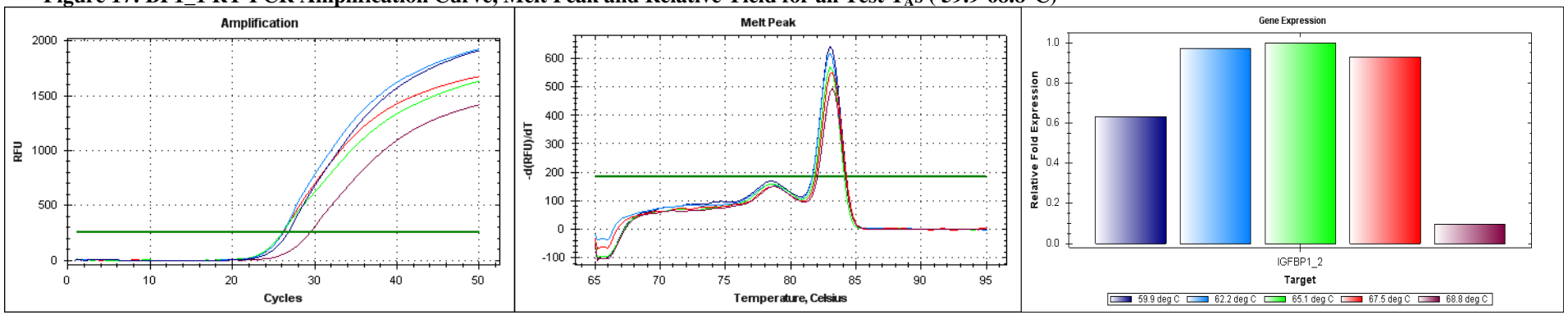


Figure 18: BP1_2 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

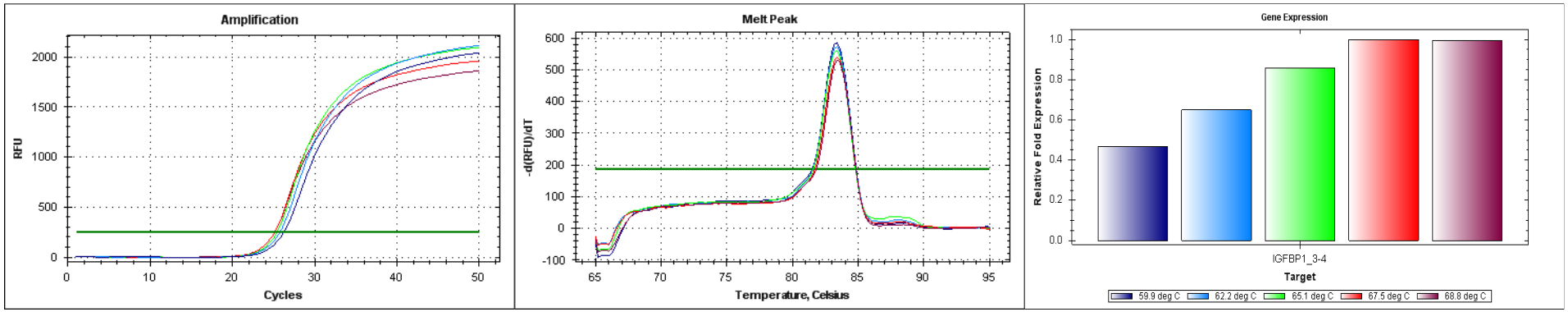


Figure 19: BP1_3/4 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

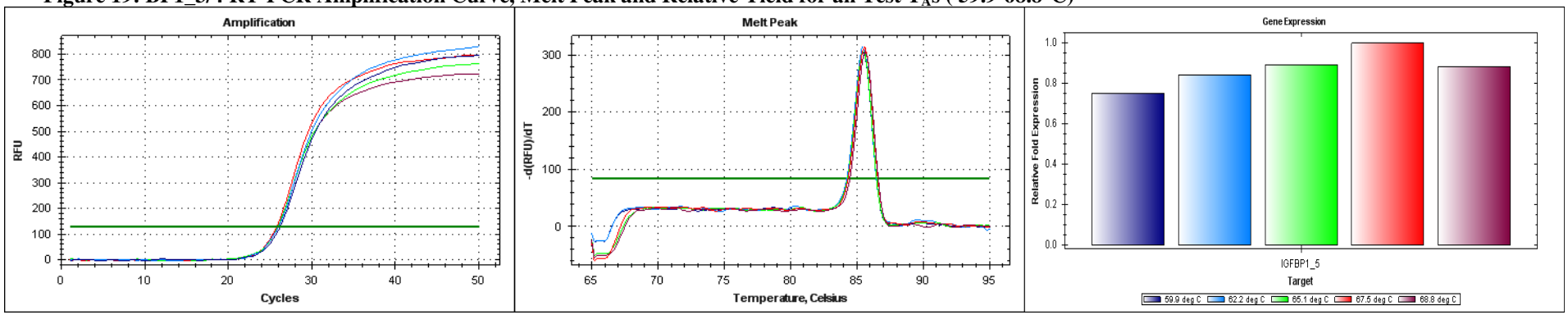


Figure 20: BP1_5 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

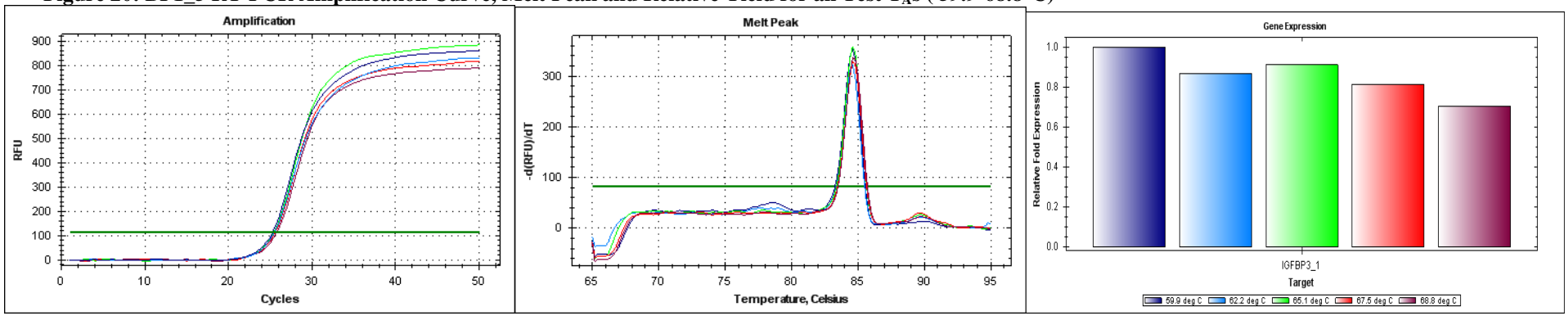


Figure 21: BP3_1 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

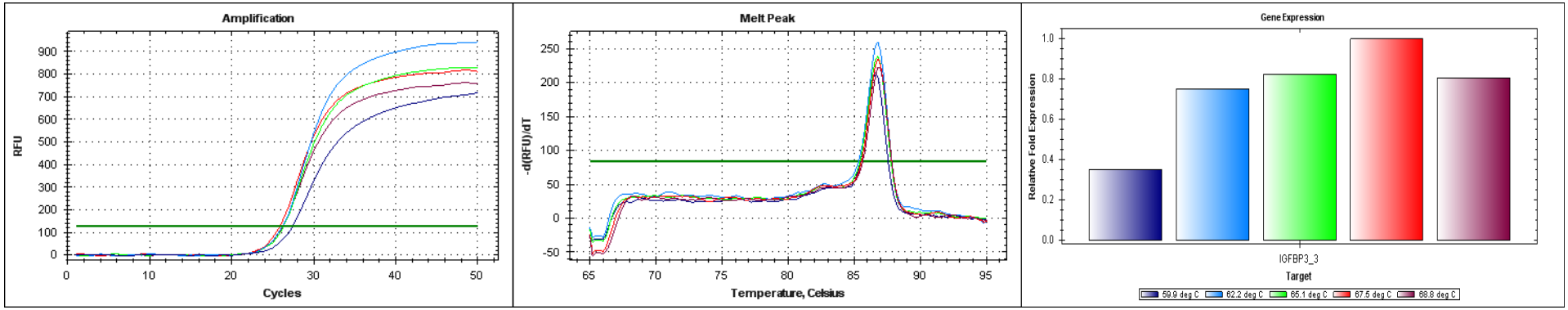


Figure 22: BP3_3 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

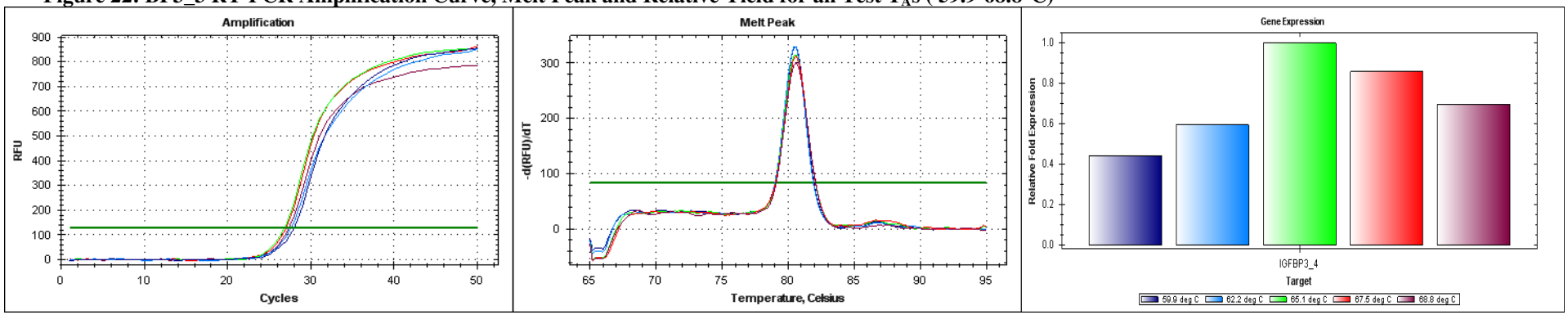


Figure 23: BP3_4 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

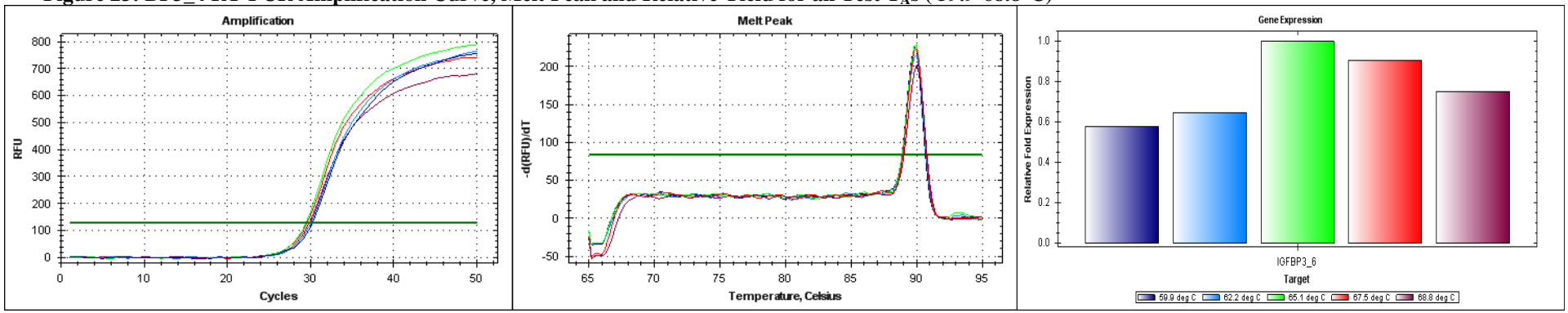


Figure 24: BP3_6 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

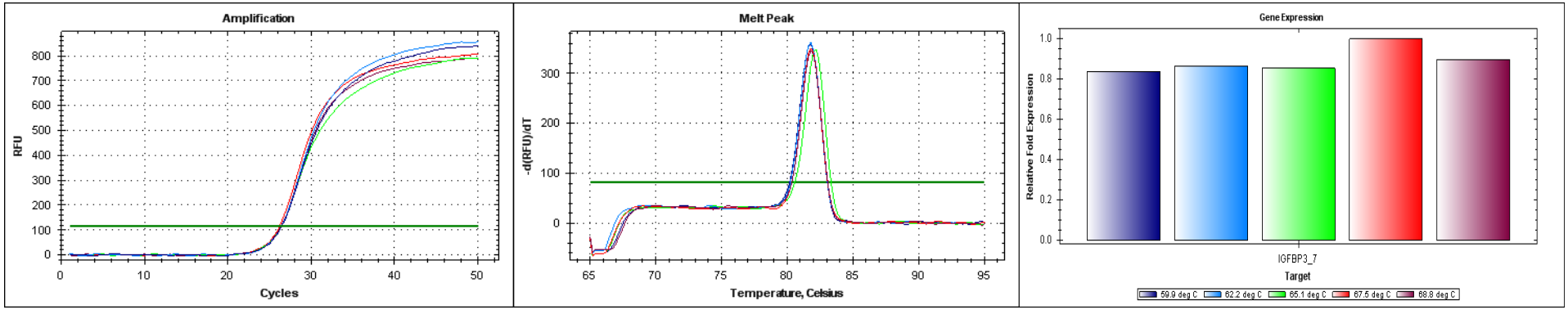


Figure 25: BP3_7 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

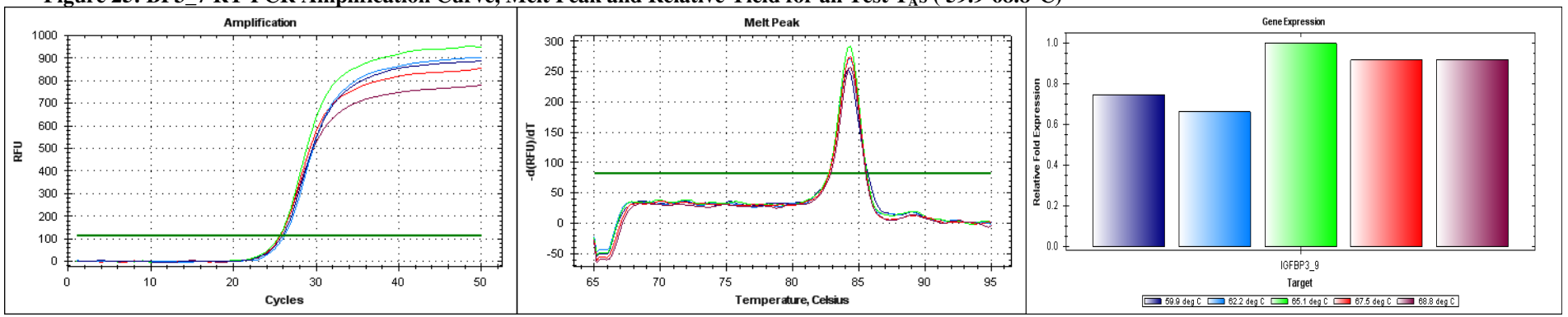


Figure 26: BP3_9 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

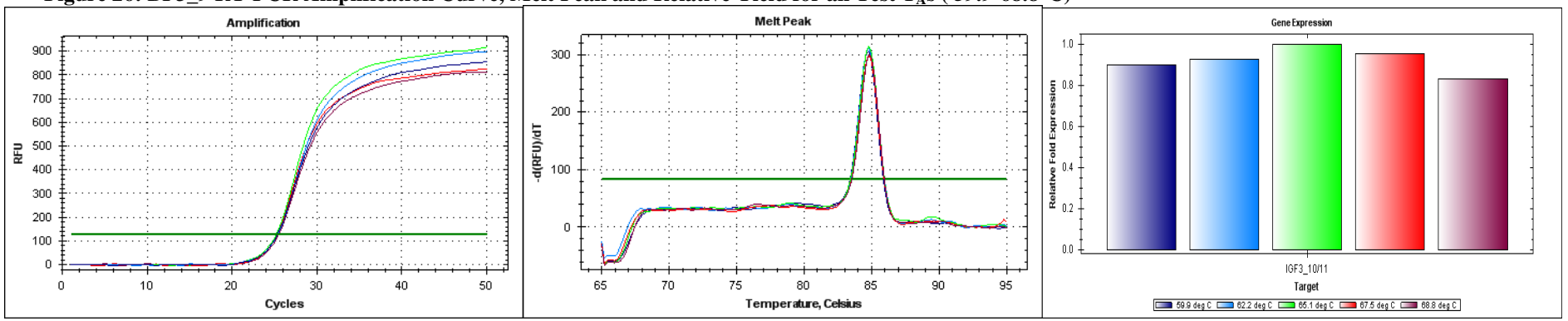


Figure 27: BP3_10/11 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

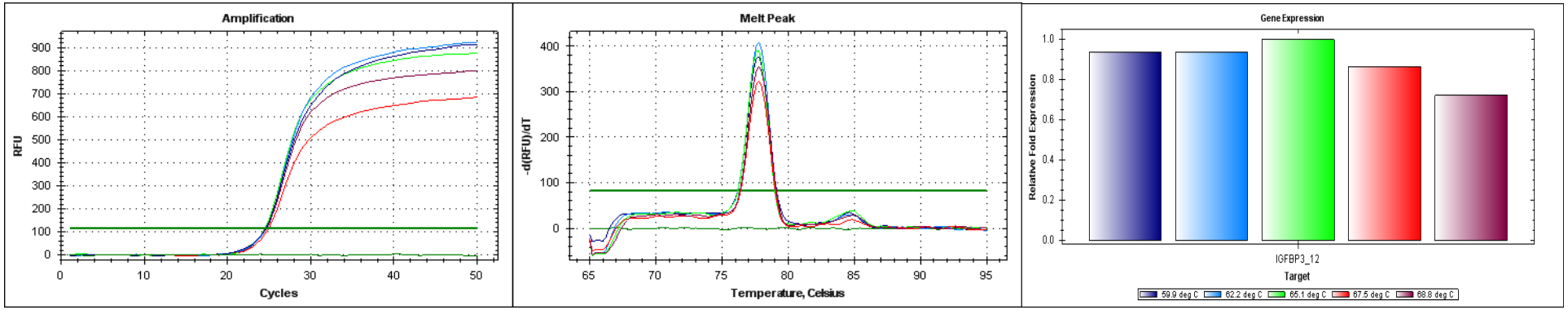


Figure 28: BP3_12 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

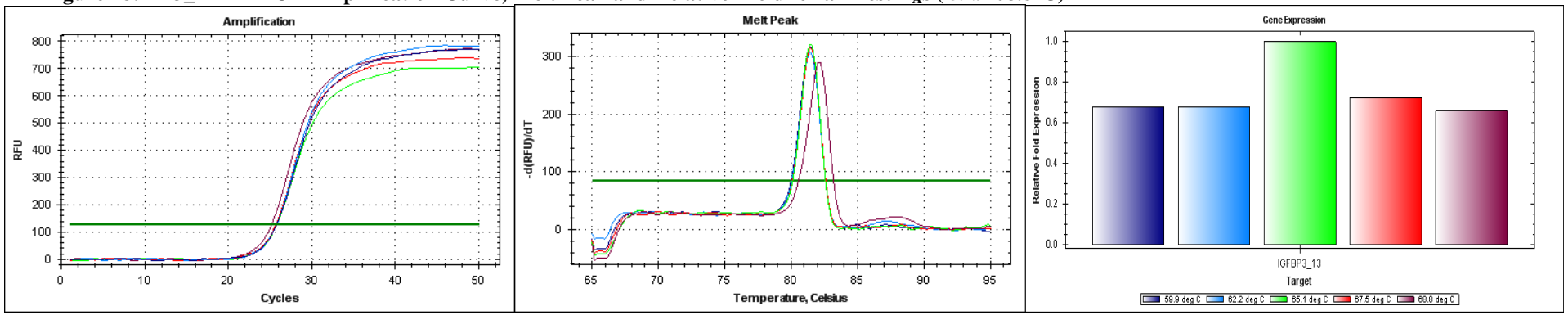


Figure 29: BP3_13 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

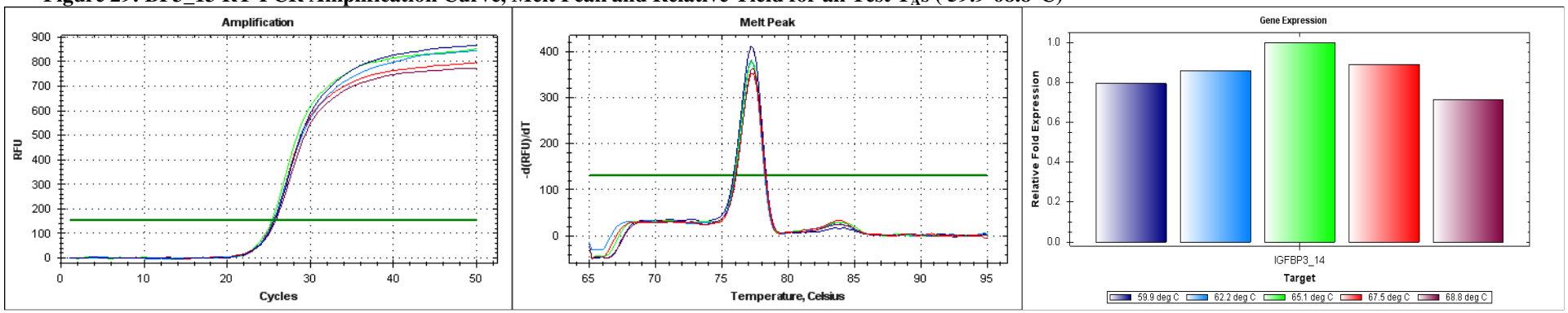


Figure 30: BP3_14 RT-PCR Amplification Curve, Melt Peak and Relative Yield for all Test T_As (59.9-68.8°C)

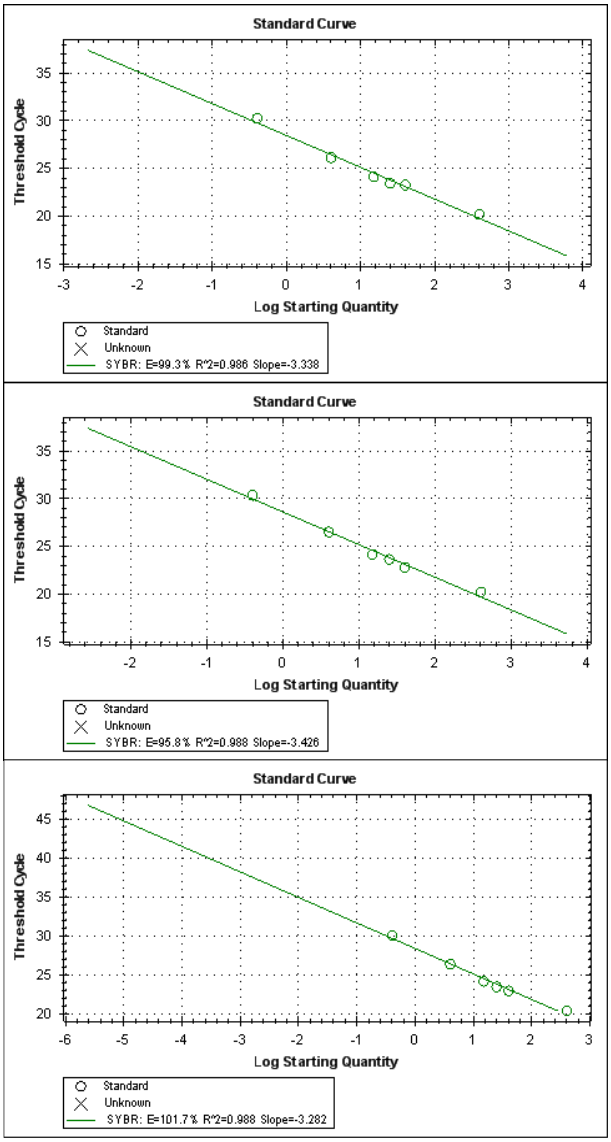


Figure 31: IGF1_1, IGF1_2 and IGF1_3 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

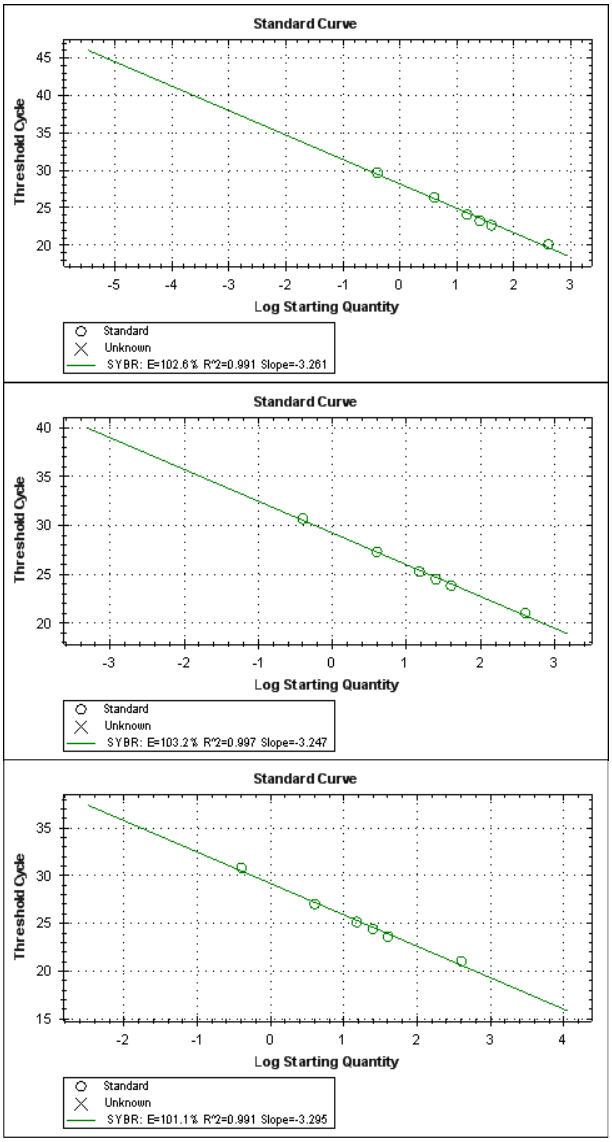


Figure 32: IGF1_4 IGF1_5 and IGF1_6 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

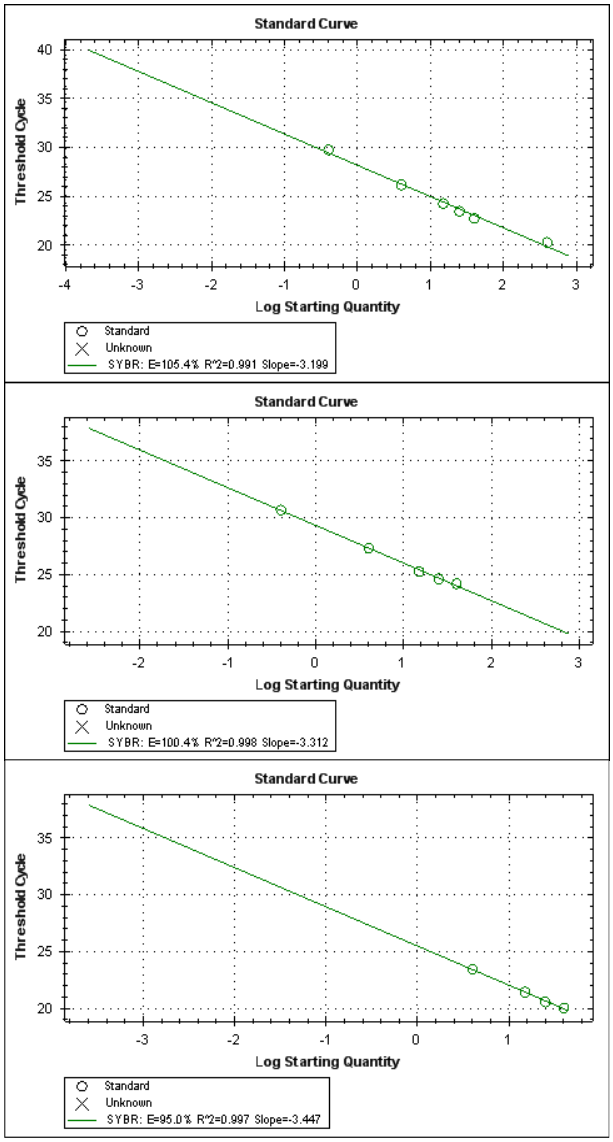


Figure 33: IGF1_7 IGF1_8 and IGF1_9 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

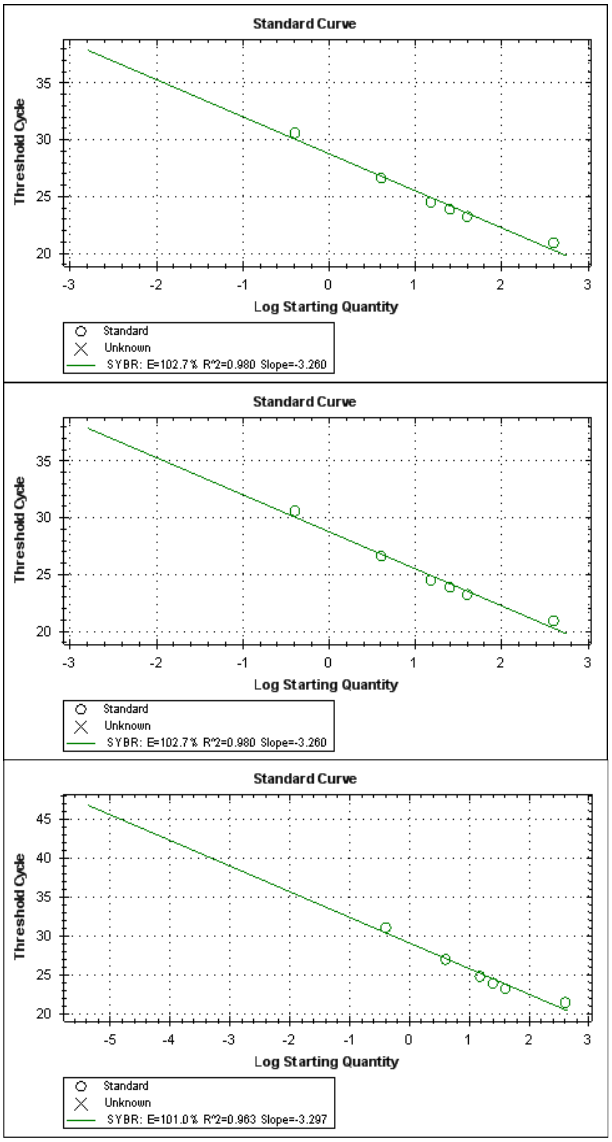


Figure 34: IGF1_10, IGF1_11 and IGF1_12 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

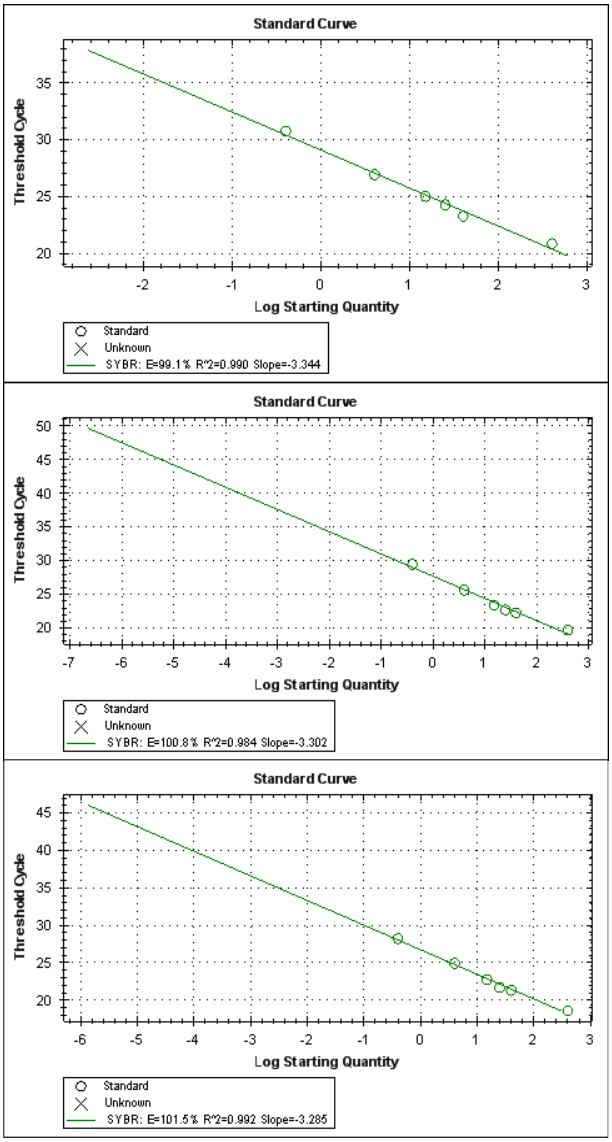


Figure 35: IGF1_13, BP1_1 and BP1_2 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

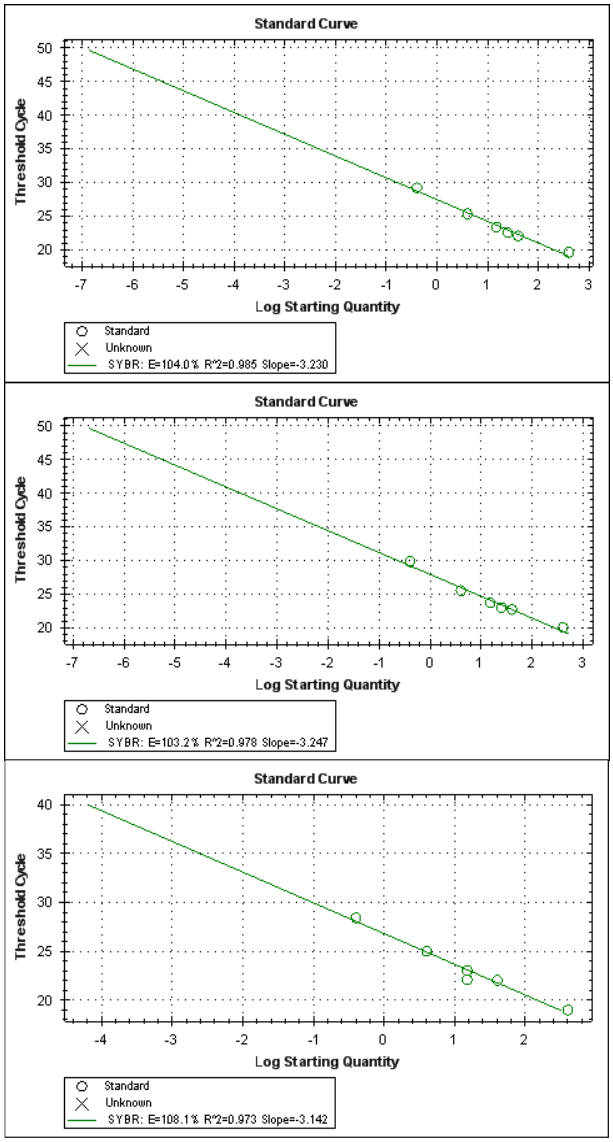


Figure 36: BP1_3/4, BP1_5 and BP3_1 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

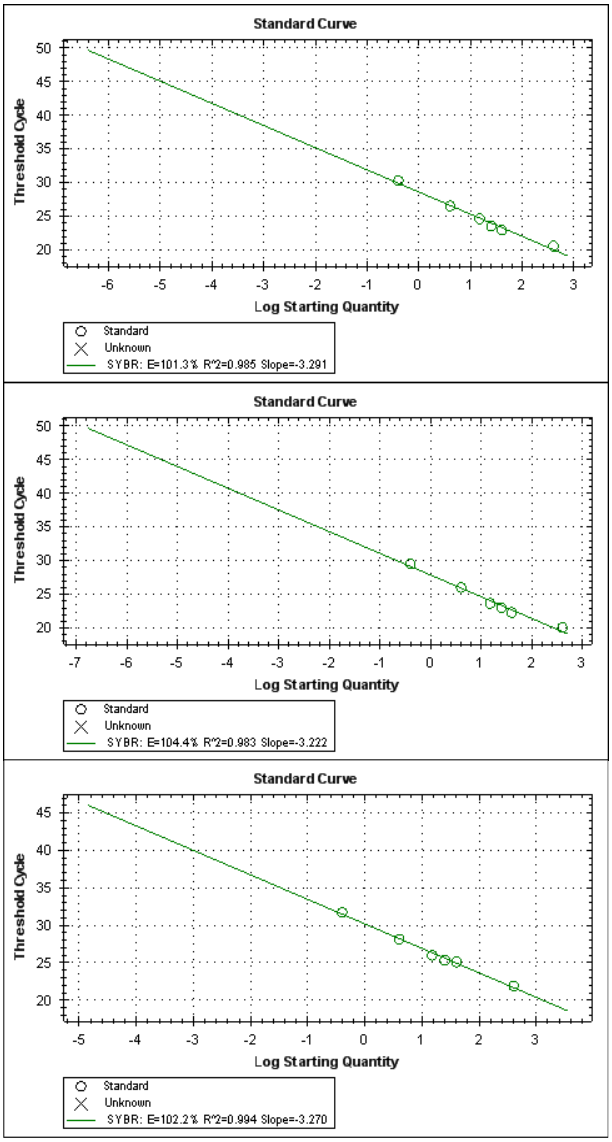


Figure 37: BP3_3, BP3_4 and BP3_6 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

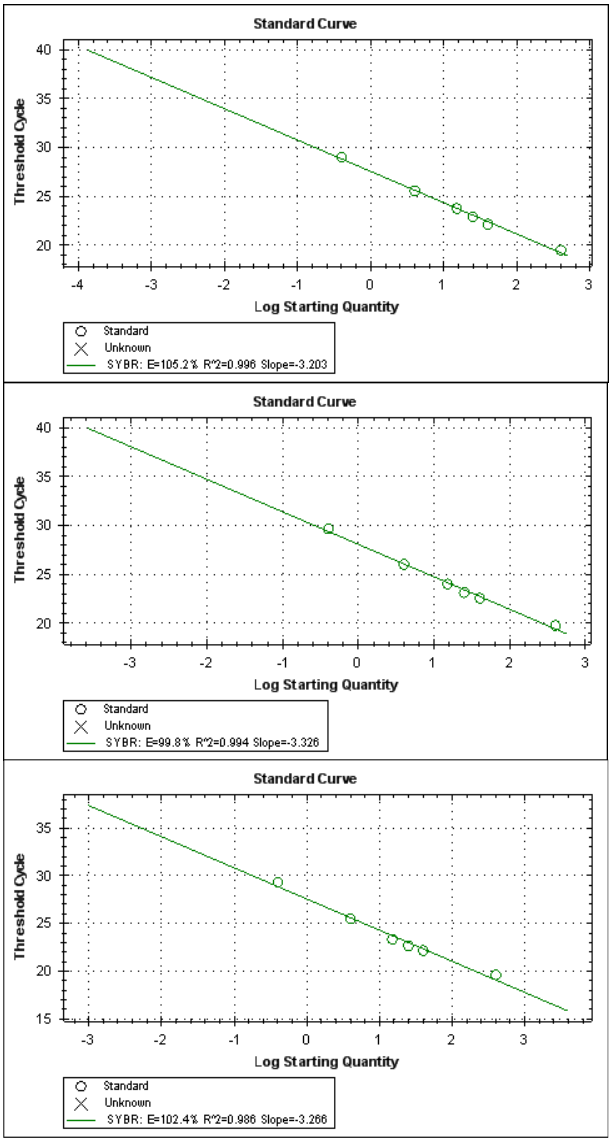


Figure 38: BP3_7, BP3_9 and BP3_10/11 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)

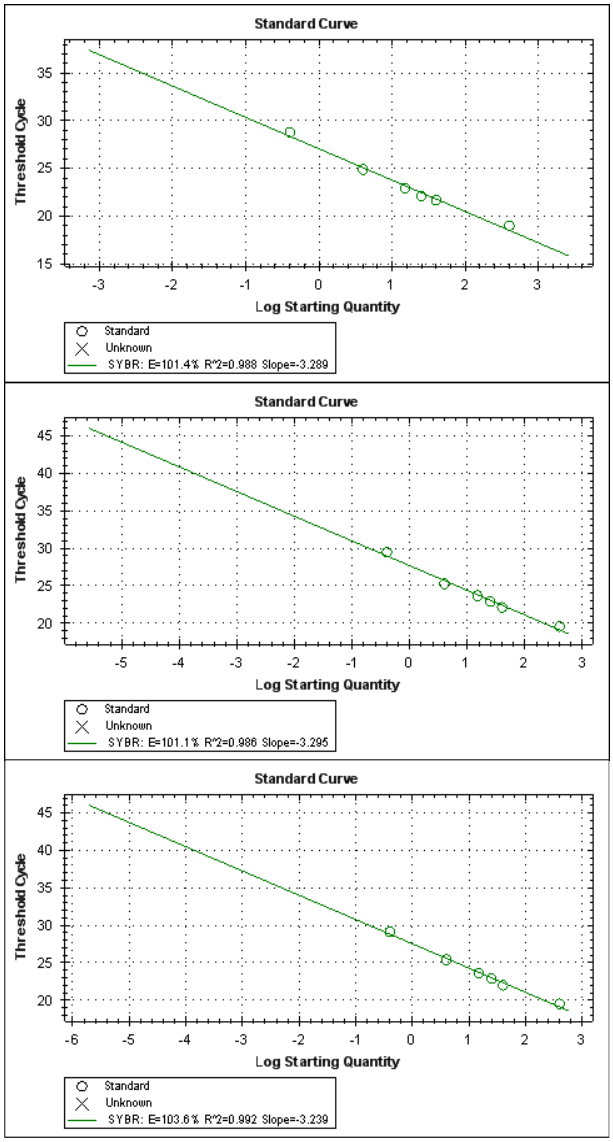


Figure 39: BP3_12, BP3_13 and BP3_14 RT-PCR Standard Curves for Template ($6x10^{1}$ – $6x10^{-3}$ fMoles)