WESTERN SYDNEY
UNIVERSITY

300820

GENES, GENOMICS AND HUMAN HEALTH

School of Science

2021 On-campus practical classes

Introduction to Genome Analysis

Laboratory Rules

Laboratory staff will refuse entry to any student who does not follow these rules.

Attendance

- As computer and practical classes contribute a total of 40% of the overall assessment for this unit (consisting of a single final report) students are strongly encouraged to attend all classes
- **Students who are more than 10 minutes late will not be admitted to the class**

Laboratory Notebooks

- All students must have a laboratory notebook
- While laboratory notebooks will not be marked, students are expected to maintain an accurate and detailed record of all experiments and results as these records will be important in completing the final assignment.
- You will be expected to prepare protocols (written, or in the form of flow charts) for each practical class, and to include detailed descriptions of experimental procedures and analysis in their laboratory notebooks
- **If you do not prepare your laboratory notebook prior to coming to class, do not expect assistance from laboratory staff if you do not understand the tasks in each practical class**
- You will work in groups of 2

Clothing

- You are required to bring your own lab coat
- Wearing safety goggles is compulsory
- When handling reagents, gloves must be worn
- Remove your gloves when opening doors
- Closed shoes only (heel and instep protected)
- Flat shoes, not high heels

What is expected in the Genes, Genomics, and Human Health practical classes?

- Protocols read and an experimental protocol written into your laboratory notebook BEFORE the practical commences
- When you are in the laboratory, discuss with your partner(s) how the experiments will be done and assign tasks and responsibilities
- Include enough detail in your laboratory notebook so that someone else reading your notes would understand what you have done and be able to repeat the experiments
- A laboratory notebook is a written record of your experiments. It is not a work of art! We all make mistakes, change protocols and adjust volumes- a laboratory notebook should include all of this

Genes, Genomics and Human Health

On-campus Laboratory Classes: rationale and aims

Genome analysis can involve analysis of an entire genome (genome-wide analysis) or the analysis of a specific gene or locus (targeted analysis).

Over three on-campus practical classes you will:

1. Understand the principles of library construction for whole-genome DNA sequencing
2. Carry out targeted copy-number analysis of a gene that is often amplified in lung cancer cells

The aims of this part of the course are:

1. Learn to cooperate and work effectively in small groups
2. Using DNA methylation-sensitive restriction endonucleases and ligation-mediated PCR, carry out the first steps for constructing a library for next-generation DNA sequencing
3. Using TaqMan quantitative PCR (qPCR) technology, compare the copy number of the epidermal growth factor gene (EGFR) locus between normal and cancerous cells
4. Analyse the results of copy-number qPCR using basis statistical methods
5. Write a laboratory report that demonstrates your understanding of the method and the application of these techniques in human genome analysis

GGHH 300820 On-campus Practical Class Schedule 2021

Week	Group	Wednesday (Parramatta)	Friday (Campbelltown)
1 1/3	Online practical classes		
2 8/3			
3 15/3	Groups 01 and 02	Parramatta, PS-EHa.G.27 GROUP 01: Practical Lab 1, 9am-12pm GROUP 02: Practical Lab 1, 1-4pm	Campbelltown, CA- CA-8.G.07 GROUP 01: Practical Lab 1, 9am-12pm GROUP 02: Practical Lab 1, 1-4pm
4 22/3		Parramatta, PS-EHa.G.27 GROUP 01: Practical Lab 1, 9am-12pm GROUP 02: Practical Lab 1, 1-4pm	Campbelltown, CA- CA-8.G.07 GROUP 01: Practical Lab 2, 9am-12pm GROUP 02: Practical Lab 2, 1-4pm
5 29/3		Parramatta, PS-EHa.G.27 GROUP 01: Practical Lab 1, 9am-12pm GROUP 02: Practical Lab 1, 1-4pm	Good Friday 2nd April
6 5/4	Groups 03 and 04	Parramatta, PS-EHa.G.27 GROUP 03: Practical Lab 1, 9am-12pm GROUP 04: Practical Lab 1, 1-4pm	Campbelltown, CA- CA-8.G.07 GROUP 01: Practical Lab 3, 9am-12pm GROUP 02: Practical Lab 3, 1-4pm
7 12/4		Parramatta, PS-EHa.G.27 GROUP 03: Practical Lab 1, 9am-12pm GROUP 04: Practical Lab 1, 1-4pm	Campbelltown, CA- CA-8.G.07 GROUP 03: Practical Lab 1, 1-4pm
8 19/4		Parramatta, PS-EHa.G.27 GROUP 03: Practical Lab 1, 9am-12pm GROUP 04: Practical Lab 1, 1-4pm	Campbelltown, CA- CA-8.G.07 GROUP 03: Practical Lab 2, 1-4pm
9 26/4		**Mid-session break**	
			Campbelltown, CA- CA-8.G.07 GROUP 03: Practical Lab 3, 1-4pm

Laboratory Notebook

Why is a laboratory notebook important?

A laboratory notebook is a record of your aims, methods and results. It is a complete record of everything you have done in the laboratory. It is essential for writing up results for reports and scientific publications. In the case of research that may lead to a new patent and intellectual property issues, it is a legal record of your discoveries.

You have one laboratory notebook and one only – **do NOT use loose pieces of paper or a second book for calculations**. EVERYTHING you do in the lab is entered in FULL DETAIL into your laboratory notebook. It does not matter if your first calculations were wrong, or a protocol needed to be amended – this is part of the learning process and is a normal and expected part of the scientific method.

If you are not prepared when you come to class – if you have simply printed out the laboratory manual or you are working from loose pieces of paper – **the laboratory demonstrators and supervisor will NOT answer any of your questions** until you have completed the laboratory notebook.

Overview of the Practical Classes: Constructing a Library for Next-Generation DNA Sequencing and Idetifying Copy Number Variation in Cancer Cells

Constructing a Library for Next-Generation DNA: the *Hpa*II tiny fragment Enrichment by Ligation-mediated PCR (HELP) assay

Figure 1. Simplified workflow of Next-Generation DNA sequencing. In the practical classes you will be provided with purified genomic DNA (gDNA); you will fragment the DNA in 2 separate reactions using two different restriction endonucleases called *Msp*I and *Hpa*II that can be used for analysing DNA methylation. The library will be prepared by ligating DNA sequencing adapters to the ends of the fragmented DNA: these adaptors are essential for amplification of the library using PCR and for hybridization to DNA sequencing primers. You will follow the workflow to the point of amplifying the library.

Epigenetics is a broad term that links phenotypic changes in cells and organisms to reversible chemical changes to our DNA and to the proteins that bind our DNA. DNA methylation is an example of epigenetic modification of the genome and the HELP assay is a simple assay that examines genome-wide DNA methylation. Epigenetic modifications are key regulators of gene expression, but also have other effects on a genome. There are a number of events that can change the epigenetic modification of a gene, including DNA mutations and cellular differentiation.

DNA methylation occurs across the genome, with methylation occurring at the 5' cytosine base (C) in the dinucleotide 5'-CG-3' and occurs on both strands of DNA. In some regions of the genome, and most typically at promoters of a specific class of genes, 5'-CG-3' dinucleotides can extend over tens or thousands of bases and are called CpG islands. The methylation of DNA is a repressive mark, inhibiting gene transcription (***see online Lecture 5***).

The HELP Assay makes of the two restriction endonucleases that are isoschizomers. Isoschizomers are different restriction endonucleases that recognise the same DNA sequence. The restriction endonucleases *Msp*I and *Hpa*II recognise the DNA sequence 5'-CCGG-3' but exhibit a key difference in their digestion of DNA: methylation of the second cytosine residue inhibits *Hpa*II activity but not *Msp*I. This difference in activity has resulted in a number of simple assays to detect differences in DNA methylation.

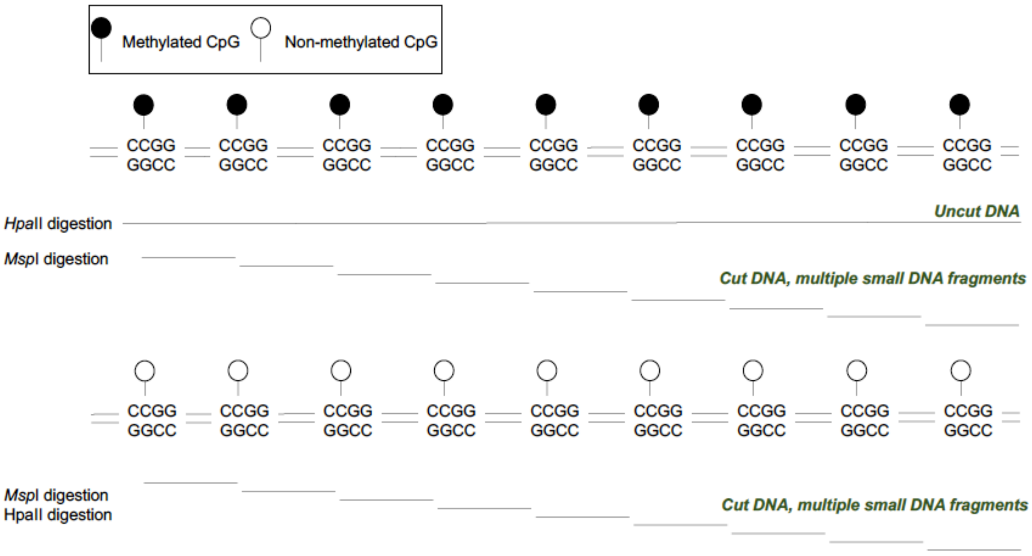


Figure 2. Basis of the HELP assay and the digestion of methylated and unmethylated DNA by *Msp*I or *Hpa*II. In this simplified example, there are multiple recognition sequences for *Hpa*II or *Msp*I (5'-CC-3'). A methylate cytosine is indicated by the black closed circle, unmethylated cytosine is indicated by the open white circle. When DNA is methylated, *Hpa*II cannot cut the DNA, while the DNA is cut into multiple small fragments by *Msp*I. In contrast, when DNA is not methylated, both *Hpa*II and *Msp*I cut the DNA, with each producing multiple short DNA fragments. The HELP assay amplifies and sequences these multiple short DNA fragments.

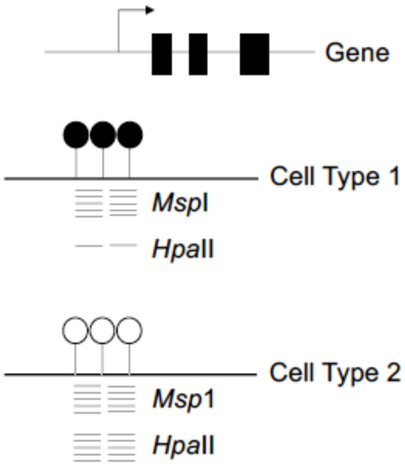


Figure 3 Differences in DNA methylation between two cells will result in a difference in the abundance of *Hpa*II-generated DNA fragments. The above figure shows a simple example of two cell types and a difference in DNA methylation at the promoter of a gene between those two cell types. High methylation will result in fewer *Hpa*II fragments (remember this is a population of cells and there will be variation in methylation within the population). Next generation DNA sequencing can reveal the frequency of short *Hpa*II-dependent fragments (normalised to *Msp*I-dependent fragments), with a difference in frequency between two cell types being correlated with a difference in DNA methylation.

Although the polymerase chain reaction has allowed scientists to specifically amplify very small amounts of DNA, the exponential increase in DNA copy number on the PCR reaction makes it difficult, if not impossible, to directly relate the amount of product generated in a PCR reaction to amount of starting material. This problem was solved with the development of technologies that use the incorporation of fluorescent dyes into the amplified DNA to measure in real-time the increase in DNA product through an increase in fluorescent dye incorporated into the amplified DNA. The simplest assay uses the dye SYBR green, which intercalates (inserts itself into) double stranded DNA in the PCR reaction. Importantly, SYBR green only emits a fluorescent signal when it becomes incorporated into newly synthesized DNA. Real-time PCR allows molecular biologists to quantify the amount of starting material by measuring the increase in fluorescence at each cycle of the PCR reaction. This is done using a specialized PCR machine that is fitted with an optical system capable of reading the fluorescent signal.

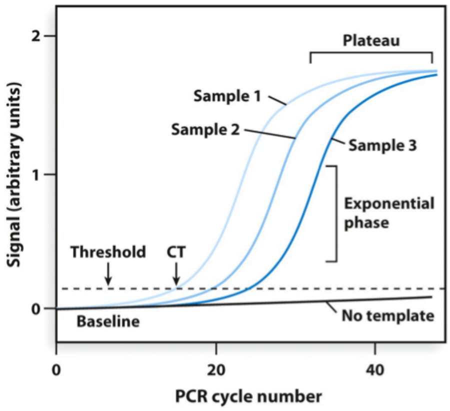


Figure 7-10
Molecular Biology: Principles and Practice
© 2012 W. H. Freeman and Company

Figure 4 Quantification of Gene Expression using Real-time PCR. The amount of amplified gene target in a PCR reaction is directly proportional to the input amount of target only during the exponential phase of PCR amplification. Critical for the quantitative analysis of a change in gene expression is a technology that can measure the product of the target gene within the exponential phase. The cycle at which fluorescence increases above a defined threshold signal is known as the threshold cycle, or CT. Identification of CT allows gene amplification in the exponential phase to be measured.

Analysis of Copy Number Variation at the EGFR gene in lung cancer using the TaqMan Assay

The Epidermal Growth Factor Receptor (EGFR) is a receptor tyrosine kinase that is important for cell division and differentiation. It is also commonly mutated in cancer cells, with approximately 10% of non-small cell lung cancers exhibiting increased copy number (amplification) of the *EGFR* gene. Amplification of EGFR leads to increased expression of EGFR at the cell surface, increased intracellular signalling, and increased proliferation.

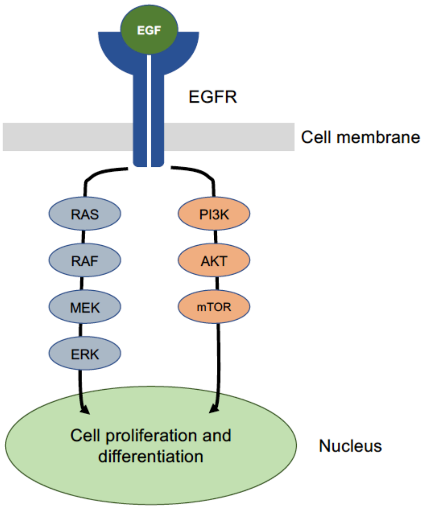


Figure 5. The Epidermal Growth Factor Receptor (EGFR) is a dimeric membrane receptor tyrosine kinase. The ligand epidermal growth factor binds the receptor, activating an intracellular kinase signalling cascade, resulting in changes in gene expression and increased cell proliferation.

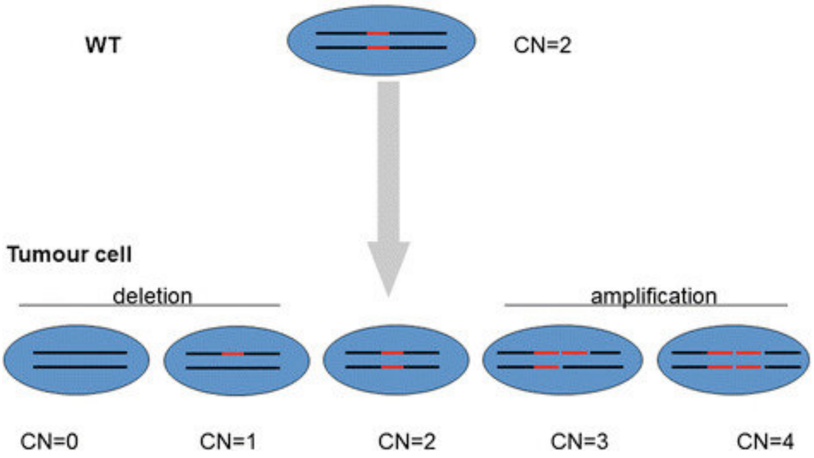
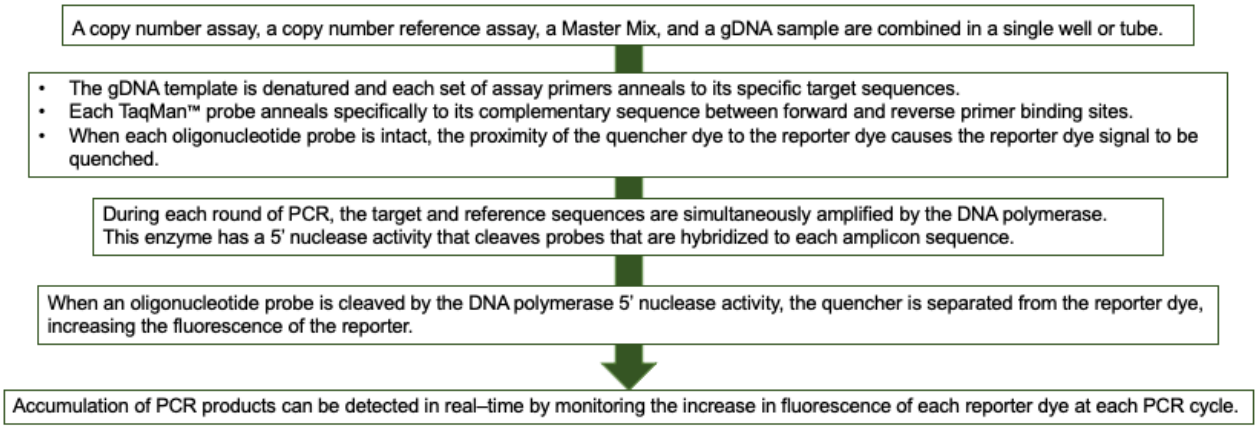


Figure 6. Cancer cells can exhibit a range of copy number variations (CNV). CNV can be localised to a specific gene locus, or it can be more extensive, involving chromosome arms or whole chromosomes (anueploid).

The Applied Biosystems TaqMan Copy Number Assays use gold-standard TaqMan DNA minor-groove binding technology (MGB) probe chemistry to evaluate the copy number of genomic DNA targets using Applied Biosystems real-time PCR instruments and software. The TaqMan *EGFR* Copy Number Assay is run together with a TaqMan Copy Number Reference Assay in a duplex qPCR reaction (a duplex reaction is one where two targets are amplified simultaneously); the copy number assay detects the target sequence, and the reference assay detects a sequence that is known to be present in two copies in the diploid genome. The results are then analyzed by the relative quantitation method using Applied Biosystems CopyCaller Software.



A copy number assay, a copy number reference assay, a Master Mix, and a gDNA sample are combined in a single well or tube.

- The gDNA template is denatured and each set of assay primers anneals to its specific target sequences.
- Each TaqMan™ probe anneals specifically to its complementary sequence between forward and reverse primer binding sites.
- When each oligonucleotide probe is intact, the proximity of the quencher dye to the reporter dye causes the reporter dye signal to be quenched.

During each round of PCR, the target and reference sequences are simultaneously amplified by the DNA polymerase. This enzyme has a 5' nuclease activity that cleaves probes that are hybridized to each amplicon sequence.

When an oligonucleotide probe is cleaved by the DNA polymerase 5' nuclease activity, the quencher is separated from the reporter dye, increasing the fluorescence of the reporter.

Accumulation of PCR products can be detected in real–time by monitoring the increase in fluorescence of each reporter dye at each PCR cycle.

Figure 7. Flow-chart of the TaqMan Copy Number Variation assay

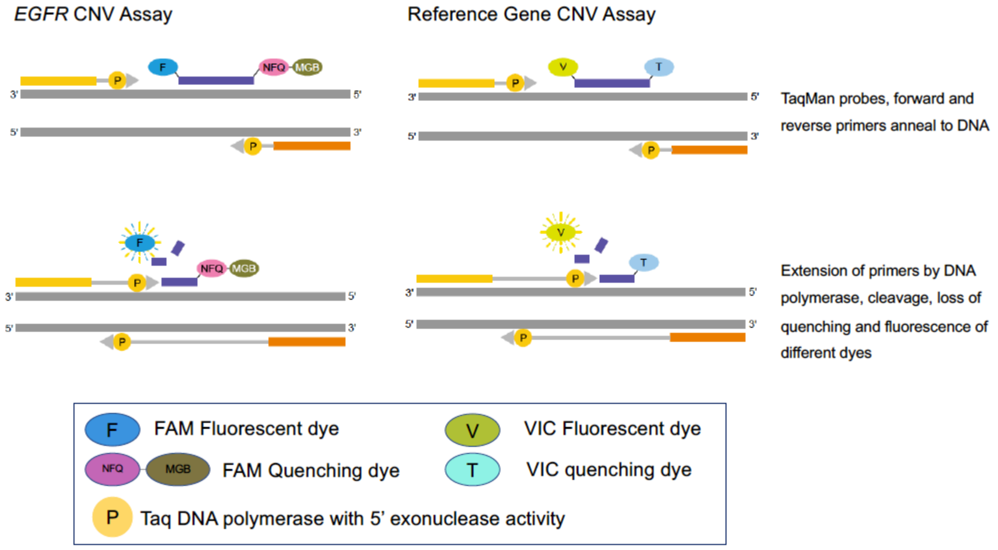


Figure 8. Details of the CNV Taqman Assay. Duplex PCR simultaneously amplifies the *EGFR* gene and the *RNaseP* reference gene, which are detected using different fluorescent dyes (FAM and VIC, respectively). Extension of the forward and reverse primers by the Taq DNA polymerase cleaves the TaqMan probe, resulting in the release of a fluorsescent signal.

More information on the TaqMan assay can also be found at:

https://www.youtube.com/watch?v=fkUDu042xic

Table 1. Schedule of activities

Practical Class	**Help Assay**	**TaqMan Assay**
1	Digest Genomic DNA with *Hpa*II and *Msp*I (2 separate reactions) and set up ligation reaction. Run agarose gel of *Msp*I and *Hpa*II digested DNA	Setup TaqMan Assay
2	PCR amplification of ligated DNA. Run agarose gel of amplified material	Receive TaqMan results, class discussion
3	Discuss results and examine DNA sequencing data (provided)	

PRACTICAL CLASSES

Face masks are compulsory and must bring your own face mask

- **DO NOT BE LATE**
- **WEAR THE CORRECT FOOTWEAR (FOOT FULLY COVERED)**
- **PLEASE PAY CAREFUL ATTENTION TO THE SAMPLES AND REAGENTS YOU ARE PROVIDED WITH**
- **MAKE NOTE OF THIS IN YOUR LABORATORY NOTEBOOK**
- **ALL TABLES MUST BE COMPLETED IN YOUR LABORATORY NOTEBOOK WITH NOTES/FLOWCHARTS DESCRIBING THE METHODOLOGY**
- **YOU MUST BRING YOUR COMPLETED LABORATORY NOTEBOOK INTO THE PRACTICAL CLASS**
- **DO NOT BRING LOOSE LEAF PRINTED LABORATORY MANUALS INTO THE PRACTICAL CLASS**

Practical Class 1

Before we start:

1. There will be a laboratory induction
2. We will have a brief discussion about the practical class

Order of activities	Help Assay	TaqMan Assay
Activity 1	Digest Genomic DNA (gDNA) with *Hpa*II and *Msp*I (2 separate reactions). *Human gDNA sourced from human blood (buffy coat)*	
Activity 2		Prepare TaqMan Assay. *Human DNA is HCC827 Purified Genomic DNA and human DNA from blood*
Activity 3	Set up ligation reaction for *Hpa*II and *Msp*I-digested samples.	
Activity 4	Run agarose gel of *Msp*I and *Hpa*II digested gDNA	

Activity 1: Digestion of Genomic DNA with HpaII or MspI

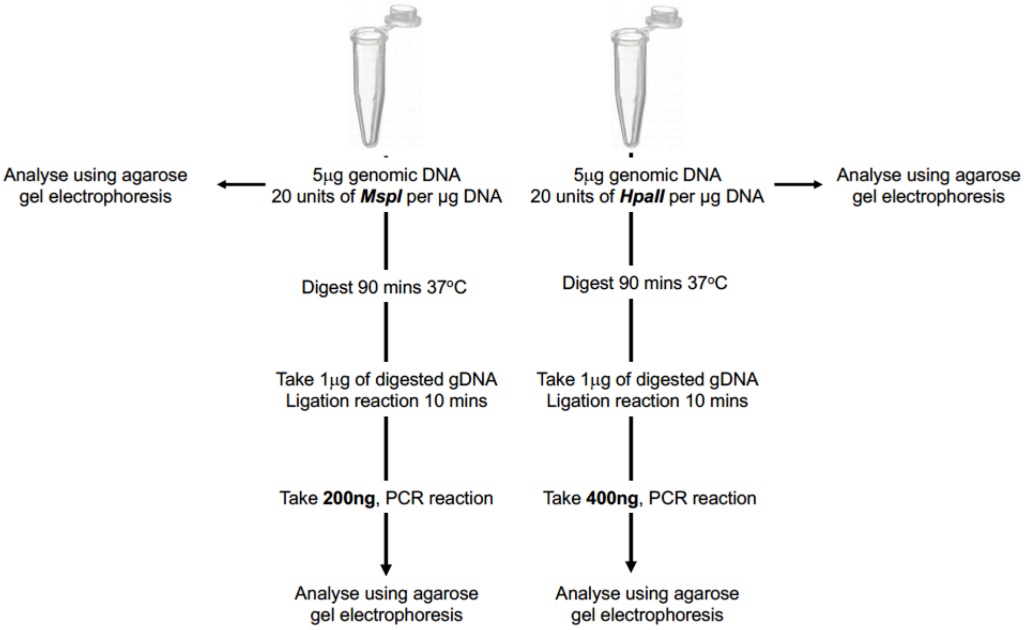


Figure 9. Diagram showing the setup of digestion of human DNA with *Msp*I or *Hpa*II.

Note: in each class several students will be asked to set up 6 tubes in total for gDNA with 3 tubes for MspI and 3 tubes for HpaII. Digested DNA will then be distributed throughout the class for the ligation and PCR assay.

Table 2 Restriction endonuclease digest of genomic DNA

Reaction components	**Details**	**Amount in final reaction**	**Volume in final reaction (µl)**
gDNA	200 ng/µl	5µg	
HpaII **OR** MspI *	*Msp*I: 20 units/µl *Hpa*II: 10 units/µl	20 units	
Restriction endonuclease buffer	Provided as 10x stock	1x in final reaction	
Water	Add to make final reaction volume 50µl	Add to make final reaction volume 50µl	
DNA digestion	1hr, 37°C	-	-

**** Remember that you need to set up one tube for MspI and 1 tube for HpaII, see Figure 9***

Activity 2: Prepare TaqMan Assay.

This activity uses Human genomic DNA from blood
and HCC827 purified genomic DNA

In this activity you will **work in pairs**. Each pair will receive 3 DNA samples, each at a concentration of 5ng/μl:

1. Undiluted human DNA from blood
2. a 1:1 mixture of human DNA from blood with HCC827 DNA
3. Undiluted HCC827 DNA

Using pooled class results we will undertake a statistical analysis to determine if there is a significant difference in EGFR copy number between each sample.

Procedure (see following page for a flow chart of the procedure):

Always use clean pipette tips, and the correct pipette, for each addition

Step 1: You are provided with 3 samples of gDNA, each at concentration of 5ng/μl (described above). You will prepare 3 x 0.2ml reaction tubes by adding 20ng of DNA from each sample to a corresponding reaction tube

Step 2: Using a single 1.5ml reaction tube, you will prepare the mixture shown in Table 3

Step 3: You will then add 16μl of the completed mixture from Table 3 to each of the 3 reaction tubes that contain the gDNA

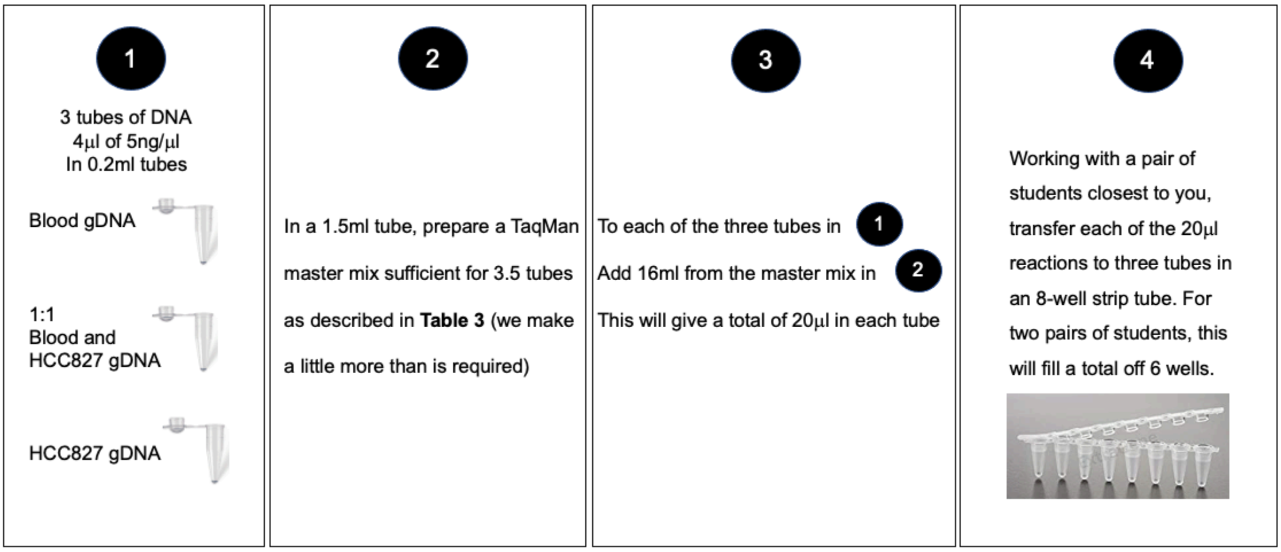
Step 4: Working with another pair of students, transfer the contents of tubes 1-3 to three wells of an 8-well strip tube. ***DO NOT WRITE ON THE LIDS OF THE 8-WELL STRIP TUBES .*** The 8-well strip tube **MUST** be loaded as follows:

Tube number	1	2	3	4	5	6	7	8
TaqMan Reaction	Pair 1			Pair 2			EMPTY	EMPTY
	1	2	3	1	2	3		

Table 3 TaqMan Reaction

Reaction components	Details	Concentration in final reaction	Volume in final reaction (μl) for a single tube	Volumes for 3.5 tubes
TaqMan Master Mix	Provided as a 2X* concentration	1X		
TaqMan EGFR Copy Number Assay	Provided as a 20X concentration	1X		
TaqMan RNaseP (reference) Copy Number Assay	Provided as a 20X concentration	1X		
Nuclease-free water	-	Add to make final total reaction volume 16μl for a single tube	Add to make final total reaction volume 16μl for a single tube	

** the X should be read as "times" – 2X is "2-times concentrated". In most cases, concentrated solutions need to be diluted to a final concentration of 1X – "1-times concentrated"*



①

3 tubes of DNA
4μl of 5ng/μl
In 0.2ml tubes

②

In a 1.5ml tube, prepare a TaqMan master mix sufficient for 3.5 tubes as described in **Table 3** (we make a little more than is required)

③

To each of the three tubes in ①
Add 16ml from the master mix in ②
This will give a total of 20μl in each tube

④

Working with a pair of students closest to you, transfer each of the 20μl reactions to three tubes in an 8-well strip tube. For two pairs of students, this will fill a total off 6 wells.

Figure 10. Flow chart for preparing the TaqMan copy number assay.

Table 4. PCR conditions for TaqMan Copy Number Assay

Step	Temperature (°C)	Time	Cycles
Initial denaturation and enzyme activation	95	10 minutes	1
Denature	95	15 seconds	40
Anneal and extend	60	60 seconds	

Note: when setting up the 7500FAST Real-time PCR machine, the reporter/quencher dyes for the EGFR and Reference (RNaseP) genes are FAM/MGB-NFQ and VIC/TAMRA, respectively.

Activity 3: Set up ligation reaction for *Hpa*II and *Msp*I-digested samples.

You will ligate the sequencing adaptor to the gDNA samples digested with either *Msp*I or *Hpa*II. The ligated samples will then be amplified this ligated fragment using the 20-mer primer 5'- `CGACGTCGACTATCCATGAACAGCGC-3'`.

```
5'- CGACGTCGACTATCCATGAACAGC -3'
                 5'- GTACTTGTCGGC-3'
```

Figure. 11 Double stranded Ligation Sequencing Adaptor. This double stranded adaptor was prepared by annealing the AdaptorSEQ_FWD and AdaptorSEQ_REV primers.

Procedure:

Step 1: The double-stranded adaptor will be provided at a concentration of 5μM

Step 2: each ligation will be done in a final total volume of 20μl with 200ng of *Hpa*II or *Msp*I-digested DNA

Step 3: Prior to adding the T4 DNA ligase, the double-stranded sequencing adaptor must be hybridized to the digested DNA. This is done by using a PCR thermocycler to heat the sample to 55°C and then cool it to 22°C over 1 hour

Step 4: Add 1μl of T4 DNA ligase and incubate for 60 minutes at 22°C

Table 5 Ligation of Sequencing adaptor to digested genomic DNA

Components	**Amount**	**Volume (to 20μl)**
*Msp*I or *Hpa*II-digested gDNA	200ng	
Sequencing Adaptor (5μM)	You will add adaptor to a final concentration of either 0.5 or 1μM – please ask the laboratory supervisor	
5X Rapid ligation buffer	1X in final reaction	
Water to a final volume of 19μl	-	
Before adding the T4 DNA ligase, heat the sample to 55°C and cool to 22°C over 1 hour		
T4 DNA ligase 5units/μl	5 units in the final reaction	

Activity 4: Run agarose gel of *Msp*I and *Hpa*II digested gDNA

Using the remaining non-ligated *Hpa*II and *Msp*I-digested gDNA, laboratory staff will assist in loading and running the samples using the E-gel system.

Practical Class 2

Order of activities	Help Assay	TaqMan Assay
Activity 1	PCR amplification of ligated DNA	
Activity 2		Receive TaqMan results, class discussion (please bring laptops)

Activity 1: PCR amplification of ligated DNA

Procedure:

Step 1: Using the primer AdaptorSEQ_FWD you will amplify the ligated gDNA

Step 2: Based on each ligation reaction containing 200ng of template DNA, the PCR amplification will include 20ng of ligated-*Msp*I-digested and 40ng of ligated-*Hpa*II-digested template

Table 6. PCR amplification of ligated gDNA

Components for PCR	**Stock concentrations**	**Amount or final concentration in reaction - *HpaII***	**Amount or final concentration in reaction - *MspI***	**μl in a reaction volume of 25μl**	
				MspI	***HpaII***
Digested and Ligated Template DNA					
MyTaq	2X	1X	1X		
AdaptorSEQ_FWD	5μM	0.5μM	0.5μM		
Nuclease-free water	-	-	-		

Table 7. PCR cycling conditions for amplification of ligated DNA

Step	**Temperature (°C)**	**Time**
Initial extension	72	10 min
Initial Denaturation	95	1 min
20 cycles	95	30 secs
	72	3 mins
Final Extension	72	10 mins
Hold	10	-

Activity 2: Receive TaqMan results, class discussion

We will examine all results and undertake a statistical analysis of the results.

Practical Class 3

Order of activities	Help Assay	TaqMan Assay
Activity 1	Run agarose gel of amplified material	
Activity 2	Discuss results and incorporate Next Generation DNA sequencing	
Activity 3	Discussion of laboratory report	

Activity 1: Run agarose gel of amplified HELP assay samples

Laboratory staff will assist in loading and running the amplified HELP assay samples using the E-gel system.

Activity 2: Discuss results and incorporate Next Generation DNA sequencing

We will discuss the outcome of the HELP assay and incorporate real-world data Next Generation DNA sequencing data into this part of the practical class.

Analysis of DNA Methylation using Short *Msp*I DNA Fragments

This track was produced as part of the ENCODE project. The track reports the percentage of DNA molecules that exhibit cytosine methylation at specific CpG dinucleotides. In general, DNA methylation within a gene's promoter is associated with gene silencing and DNA methylation within the exons and introns of a gene is associated with gene expression. Proper regulation of DNA methylation is essential during development and aberrant DNA methylation is a hallmark of cancer. In the practical class we performed the first steps in the construction of libraries that can then be sequenced using Next Generation DNA Sequencing.

We will now plot the results of an analysis of DNA methylation that was also based on the use of MspI and the generation of short DNA fragments. We will analyse the data using a software package that can be used online or downloaded onto your own computer (**Integrative Genomics Viewer, http://software.broadinstitute.org/software/igv/**).

In the data you will be analysing, genomic DNA was digested with the methyl-insensitive restriction enzyme *Msp*I and then small genomic DNA fragments were purified by gel electrophoresis and used to construct a Next Generation DNA Sequencing library. The library fragments were treated with a chemical called sodium bisulfite that will modify every non-methylated Cytosine: subsequent amplification by PCR converts every unmethylated cytosine to a thymidine leaving methylated cytosines intact. The library was sequenced, and fragments were aligned to a customized reference genome sequence. For each assayed CpG, the number of sequencing reads covering that CpG and the percentage of those reads that were methylated (methylated cytosine residues were protected from modification and are read as cytosine in the sequencing reaction). The sequencing file was processed into a format that can be read by the IGV software (a so-call BED file). Methylation status is then using the following convention:

red = 100% of molecules sequenced are methylated
yellow = 50% of molecules sequenced are methylated
green = 0% of molecules sequenced are methylated

The score in this track reports the number of sequencing reads obtained for each CpG, which is often called 'coverage'. The score is capped at 1000, so any CpGs that were covered by more than 1000 sequencing reads have a score of 1000. The BED files available for download contain two extra columns: one with the uncapped coverage (number of reads at that site) and one with the percentage of those reads that show methylation. High reproducibility was obtained, with correlation coefficients greater than 0.9 between biological replicates, when only considering CpGs represented by at least 10 sequencing reads (10X coverage, score=10). Therefore, the default view for this track is set to 10X coverage, or a score of 10.

You are provided with 4 sequencing files: 2 are from a lung cancer cell line and two are from a leukaemia cell line (T-cell). We will examine the methylation pattern in and around the genes *CDH1* and *TAL1* and the results will be incorporated into your report.

Activity 3: Discussion of laboratory report

There will be a class discussion on writing the laboratory report

Working Online with DNA

Technological advances and decreasing costs associated with DNA and RNA sequencing means it is now commonplace to report the DNA sequence of whole genomes accompanied by genome-wide gene expression data. The generation of such huge amounts of data creates an invaluable resource for all biomedical researchers.

Today, it is not uncommon for biomedical researchers to commence their projects by mining data that has been made available to researchers.

Much of the analysis done by researchers and clinicians working on the human genome involves automated processes and the generation of large datasets. Developing analytical skills that require some knowledge of statistics and programming languages such as R and python is now a pre-requisite for anyone interested in working in the field of genomics.

Some sites you may find interesting and useful:

OMIM
http://www.ncbi.nlm.nih.gov/omim/
Online Mendelian Inheritance in Man (OMIM®) is a continuously updated catalog of human genes and genetic disorders and traits, with particular focus on the molecular relationship between genetic variation and phenotypic expression. It is thus considered to be a phenotypic companion to the Human Genome Project. OMIM is a continuation of Dr. Victor A. McKusick's Mendelian Inheritance in Man, which was published through 12 editions, the last in 1998. OMIM is currently biocurated at the McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine.
Help for OMIM can be found at: http://omim.org/help/faq/#1.1

UCSC Human Genome Browser

http://genome.ucsc.edu/index.html

This website is the repository of the human genome reference sequence, as well as including many additional tracks that provide information on genome annotation, epeigenetic marks, genetic variants, DNA conservation and gene expression.

dbSNP : a database of single nuleotide polymorphisms

https://www.ncbi.nlm.nih.gov/snp

NCBI Gene: A database of information about genes

http://www.ncbi.nlm.nih.gov/gene/

NCBI Pubmed: A database of all biomedical and medical research papers and reviews

http://www.ncbi.nlm.nih.gov/pubmed/