1. Human practical: Identification of prognostic biomarkers in human prostate cancer patients

A study is being run to look for prognostic biomarkers of recurrence free survival in prostate cancer in the UK. A biomarker is a term often used to refer to measurable characteristics that reflects the severity or presence of some disease state. More generally a biomarker is anything that can be used as an indicator of a particular disease state. Biomarkers can be measured from many different things including blood, urine, saliva or tissue samples.

Study aim: To find a panel of prognostic biomarkers to predict those with poor recurrence free survival from prostate cancer to be able to offer them more aggressive treatment and spare those with a better prognosis from a treatment that is too aggressive.

Solutions

1. What would the ideal study cohort be? What factors prevent this from being the study cohort? In reality what cohort would be chosen? What issues might this cause?

    In an ideal world our cohort in this observational cohort study would be all prostate cancer patients from the UK at a particular time point. Logistically (spread over too wide an area and in too many hospitals), financially (many patients to take samples from and track) and ethically (not all patients will consent to be in the study) this would not be possible. For these reasons we need to take a sample of the patients.

    Ideally we would take a random sample of all patients from across the country with every patient having an equal chance of being included in the study. Again this would have logistical issues. Next best would be to take a random sample of hospitals across the country and take a random sample of patients within those hospitals, with the sample size being proportional to the number of prostate cancer patients and perhaps stratifying on some important variables. These two methods are most likely to give a cohort that is representative of the population of the UK therefore the results will be generalisable to the population of interest.

    Most likely is that the study cohort will be all the patients within a particular time frame who consent to be included in the study from a local hospital, or perhaps with one or two others included. Issues with the generalisability of the study findings may arise if this hospital(s) are not representative of the population of prostate cancer patients across the UK.

2. What are the pros and cons for using a retrospective or prospective cohort for this study? [A retrospective cohort study looks back over time and takes

data from patients notes and/or interviews. A prospective cohort study follows the cohort through time collecting data as it goes].

A retrospective study might not collect the correct data if it is something that is not routinely available in patient notes, they are more prone to missing data that can not be filled in at a later date. Data may not be in the correct format. Time consuming to extract data from notes. However, most of the data is already there, might have long follow up of patients, don't need to wait a long time to achieve this.

A prospective study allows collection of exactly the data that you need in the way that you need it. Less likely to have missing data, as you maybe able to chase up missing values if they are spotted quickly enough. However, they can take a long time to accrue patients and follow them up so won't get a quick result.

3. What would the outcome measure(s) be for this study? What issues can arise in the definition of the outcome measures for this study?

The main outcome is recurrence free survival. Therefore, the main outcome measure will be time to recurrence or time to death (whichever is first). However, recurrence and death need to be carefully defined. Recurrence maybe biochemical recurrence, triggered recurrence or clinical recurrence and it will be important to be clear which is the outcome. Death can be prostate cancer specific death or death from any cause.

As well as the main (or primary) outcome, we may include secondary outcomes that are additional question we wish to answer. Secondary outcomes maybe overall survival time, all cause mortality or distant metastases free survival. For which the outcome measures are time to prostate cancer specific death, time to death from any cause, time to distant metastases.

4. What factors might affect the outcome? What information may you need to collect for use in the study analysis?

- Gleason grade: pathological scoring system by which to classify prostate cancers with respect to microscopic features of cancer cells, including differentiation. Prognostic score, with higher score predicting worse outcome. Score is given from 1-5 and two scores are given to represent the two most prevalent cell types. These scores are represented as (score 1) + (score 2) and is out of 10. Gleason sum ≤ 6 is well-differentiated, 7 is moderately-differentiated and 8-10 is poorly-differentiated.
- T-stage: can be clinical, radiological or pathological. Clinical T stage is used to classify patients based on the palpable extent of disease. If it is impalpable, but found on biopsy, it is T1c. Palpable cancer is then

classified from cT2 to cT4 (with 4 being more extensive disease into the surrounding organs, i.e. rectum or bladder). Radiological is the extent of disease found on imaging (usually CT or MRI scan). Pathological cancer is classified from a prostate, which has been removed (radical prostatectomy) and is the most accurate staging, with pT3 disease representing extra-capsular spread.

- Surgical margin status (pathological report of radical prostatectomy specimen which says if a cancer is at the margin of the specimen (different pathologists report this very differently, with some using a 2mm margin, some <1mm margin, etc.) and therefore likely to have been incompletely removed. A positive margin confers a worse prognosis and many patients need further treatment, including radiotherapy, although some patients are monitored closely without further treatment.
- PSA (prostate specific antigen), serine protease enzyme present in serum, raised levels of which indicate biochemical recurrence.
- Age at diagnosis: age usually in years at the time of prostate cancer diagnosis.
- Tumour volume: amount of tumour, can be the weight of the tumour in g or percentage of the gland that is tumour.
- Treatment: surgery, chemotherapy, radiotherapy etc.
- M-Stage: is based on the radiological findings of metastases. M-score corresponds to the location of metastases from 0 (none), 1 (distant lymph nodes), 1b (bones) and 1c (other sites).
- N-Stage: the pathological findings of lymph nodes containing cancer after radical prostatectomy, when we also remove the lymph nodes in most/high risk patients, is the N-score for nodes.
- Cellularity: The proportion of a tissue sample that is tumour.

5. What are your groups? Do you need a control group in addition? If so what will your control group be?

Two groups, those who have experienced recurrence or death, those who have not. To avoid bias in the results the cohort will need to be followed up for a sufficiently long time to observe a recurrence or death, the length of follow up should be longer than the expected time to experience the outcome.

Once the panel of biomarkers has been chosen then there will also be groups for those who are test positive and test negative. A 2-way table of these two sets of groups will be constructed to carry out statistics on. The panel of biomarkers should be different between those who go on to recur or die and those who do not.

A normal group could also be included. These could be normal healthy patients, or if we are collecting tissue samples this could be normal tissue

6. What effect would a short follow-up time have? Which group will this affect more?

    The effect of having a short follow-up time will be dependent on how quickly the patients experience recurrence or death. If these events happen relatively quickly then a short follow-up time will have less of an impact. The shorter the follow-up time the less likely you are to see the events of interest. This will mean that there are patients in the study that are included in the not recurred/died group that are there because they have not been followed up long enough to experience the event. The group that will be affected most is the not recurred/died group as this will have incorrectly classified patients. The recurred/died group will have missing patients.

    We need to compare the follow-up time in the group that has not recurred/died to the time to recurrence/death in the other group. Ideally the follow-up time in the former is longer than the time to event in the latter group, therefore, the subjects will have had sufficient time to experience the event.

7. Which technologies could you use to screen for your panel of biomarkers?

    Genomics: RNA-seq, gene expression microarrays, qPCR/RT-PCR
    Proteomics: mass spectrometry, tissue microarrays

8. What factors might influence your choice of technology for this study?

    The decision would be based on your objectives, cost, logistics, feasibility and technological limitations.

9. Once your study has a panel of biomarkers from which a test for recurrence has been formulated. What is the next step?

    The test must be tested in an external validation cohort. This would be a different set of patients, perhaps in a different hospital and at a different time. The test will be tested to see how well it predicts recurrence in prostate cancer.

10. What characteristics would a clinical test formed from your biomarkers need?

    Quick result, preferably one that can be done there and then on the spot rather than being sent to a lab. A non-invasive procedure of sample collection would be best. Collecting urine and saliva is non-invasive to the patient whereas taking tissue biopsies (e.g. using needles) are invasive and

complicated. You may want to focus your efforts on secretory/soluble proteins, for example.

For a test to be adopted it needs to have a high sensitivity (proportion of recurrent cancers that the test detects) and a high specificity (proportion of non-recurrent cancers that have a negative test). Of course there is always a trade-off between sensitivity and specificity. As this test may be used to decide on a more aggressive form of treatment we would want to maximise specificity, as we would wish to spare those patients who are unlikely to experience recurrence from the harmful side effects of aggressive treatment. From the perspective of the patient it should have a high positive predictive value (probability of having prostate cancer given that you've had a positive test result) and negative predictive value (probability of not having prostate cancer when you've had a negative test result).

2. Mouse practical: Gene expression profile of wild and mutant HHEX in brain and liver development

Hematopoietically Expressed Homeobox (HHEX) is a transcription factor which, plays an important role in the proper development of brain and liver in mouse. A mutant HHEX (where all Serine and Tyrosine residues are mutated to Alanine) is hyperactive and induces fetal death in mutant homozygous mice. We are interested in identifying gene expression changes in brain and liver and in determining key pathways involved, in response to hyperactivity of HHEX gene. Note that, samples are collected from 15 day-old fetuses that are homozygous wildtype (Wt/Wt), heterozygous mutant (Wt/Mt) and homozygous mutant (Mt/Mt), which manifest distinguishing morphological characteristics.

Solutions

1) What are the scientific questions of interest in this experiment?

   a. Which genes show expression changes in response to mutant HHEX in brain and/or liver?
   b. Do homozygous and heterozygous mutant HHEX mice show similar expression profiles?
   c. Which genes show exclusive response in brain and liver respectively in response to hyperactivity of HHEX gene?
   d. Which are the pathways that are involved in HHEX hyperactivity?

2) What are you measuring?

   a. If you use RNA-seq you are measuring read counts.
   b. If you use microarray you are measuring intensities.

3) What controls samples should be included in this experiment? Why is control needed in the experiment?

   a. Wt/Wt mouse group is a control group that Wt/Mt and Mt/Mt groups will be compared against.
   b. It is difficult, if not impossible, to measure the absolute gene expression changes. In differential expression experiments, one or more control group acts as a reference to compare the groups of interest to. Both Wt/Mt and Mt/Mt groups are compared against the control (Wt/Wt group).

4) How many replicates you need to include for each group? Discuss what factors might have influence in selecting the number of replicates?

   a. Including replicates reduces variability in experimental results, increasing their significance.

b. There is no standard number of replicates recommended in an experiment. The number of replicates required, depends on various factors and therefore varies experiment to experiment. It is important to consider the number of replicates for each experiment that is carried out.

c. The absolute minimum number of replicates required is three per each experimental group. Statistical analysis cannot be carried if there are fewer than three replicates per group. The Bioinformatics Core recommend a minimum of four replicates per group to allow for analysis to be carried out should a sample drop out of a group for whatever reason.

d. In general a larger number of replicates should result in a more powerful experiment (i.e an experiment that is more likely to detect the effect size of interest should it exist in the population).

e. Number of replicates required depends on the following factors

    i. Amount of variation between experimental units. The more variability the harder it will be to detect the difference of interest therefore more replicates will be required.

    ii. Effect size of interest (difference between control and treatment). To capture small effect size of interest, one needs to use more replicates. Therefore, it is important to know what size of effect you are interested in finding within the experiment.

    iii. Statistical power required, in general the higher the number of replicates the higher the statistical power. 80% is usually taken as the amount of power to aim for in an experiment, that is there is an 80% chance of detecting a particular effect size given that it exists in the population. However, for the p-values to be repeatable 90% power is required.

    iv. The Statistical significance level to be used in the statistical tests. This is usually taken to be 0.05 but can vary in different situations and should be pre-specified.

    v. The type of alternative hypothesis (one sided or two sided alternate hypothesis)

    vi. Other factors such as cost, ethics and ability to handle the number of replicates should also be considered.

    vii. In our experiment one wants to see the effect of mutant HHEX on brain and liver development. Assuming that for each experimental group six biological replicates are required (note that both brain and liver samples can be collected from each fetus, this will give more power to any brain/liver comparisons). We need at least 18 mouse foetuses. Mouse average litter size is litter of 3–14; therefore samples may be collected from 2-6 different litters and letters should be from different male and female mouse pairs.

5) Which experimental groups will be included?

   a. There would be six experimental groups, 1) Brain from homozygous wild type mice, 2) Brain from homozygous mutant mice, 3) Brain from heterozygous mutant mice, 4) Liver from homozygous wild type mice 5) Liver from homozygous mutant mice and 6) Liver from heterozygous mutant mice.
   b. This is a 2 factorial experiment with three different levels.

	Wt/Wt	Wt/Mt	Mt/Mt
Brain	6	6	6
Liver	6	6	6

   Wt = Wild type
   Mt = Mutant

6) How will any findings be validated?

   a. Comparing your results with previous findings from literature
   b. We could re-validate the findings using experimental techniques such as qPCR, in separate samples, such samples from completely different litters. Validating on samples from different litter gives generalizability to your findings.

7) What contrasts (sample group comparisons) you make with the data?

   a. Brain Wt/Wt Vs Brain Wt/Mt for brain specific effect of heterozygous mutant
   b. Brain Wt/Wt Vs Brain Mt/Mt for brain specific effect of homozygous mutant
   c. Liver Wt/Wt Vs liver Wt/Mt for liver specific effect of heterozygous mutant
   d. Liver Wt/Wt Vs liver Mt/Mt for liver specific effect of homozygous mutant
   e. Brain Wt/Wt Vs Liver Wt/Wt, this comparison gives differentially expressed genes in brain and liver of wild type mouse.
   f. Brain Wt/Mt Vs liver Wt/Mt, this comparison gives differentially expressed genes in brain and liver of mutant heterozygous mouse. Subtracting the, Brain Wt/Wt Vs Liver Wt/Wt, gene list gives brain and liver specific effect of heterozygous mutant genotype.
   g. Like above comparison we can also find out the gives brain and liver specific effect of homozygous mutant genotype.

8) What are possible sources of bias and confounding variables in the experiment?

   a. Gender of foetus

b. Age of the mice, animal breed, diet etc
c. People, if several people involved in sample collection and extraction.
d. Litter
e. Sequencing run
f. Index
g. Cage
h. Cage position
i. Technician handling the mice
j. Extraction batch

9) How can these sources of bias and confounding be controlled?

   a. Randomisation
   b. Blinding

3. Proteomics Practical: AR interactome differences between drug responsive and resistance conditions

The androgen receptor (AR), is a type of nuclear receptor, that is activated by binding either testosterone, or dihydrotestosterone in the cytoplasm. Activated AR then translocating into the nucleus and bind to DNA to regulate the activity of various genes. Prostate cancer is associated with alterations in AR functions. In fact, drugs that prevent the production of androgens and/or block the action of the AR inhibits prostate cancer growth. However resistance to these drugs often occurs after first couple of years. In this study we are interested in identifying the AR interactome differences between drug responsive and resistance conditions.

Solutions:

1. What are the scientific questions/objectives of interest in this experiment?

   The main objective is to find out the differentially interacting AR proteins in drug responsive and resistant conditions.

2. What experimental technique/s can you use to quantify the protein interaction differences?

   Co-IPs/western blotting, Mass spectrometry based proteomics

3. What are you measuring?

   The relative difference in the same peptide between different samples. There are several ways to do this, the most rapid/sensitive being mass spectrometry based peptide analysis (Label free, SILAC, isobaric labelling).

4. Do you think controls samples should be included in this experiment? If yes, what is the role of controls in the experiment? Suggest few samples which can be used as a control.

   IgG control for determining specificity i.e. those proteins that are bound specifically to AR protein complex.

5. What are the sample groups in the experiment?

   a) AR pull-down from drug resistant and responsive cell lines/tumours
   b) IgG pull-down from drug resistant and responsive cell lines/tumours

	Drug Responsive	Drug Resistant
AR pulldown	4	4
IgG pulldown	4	4

6. What constitutes a replicate in this experiment? Are they biological or technical? How many samples/replicates should be collected? Discuss what factors might have influence in selecting the number of replicates?

Depends on the type of design.
a) Experiment with one cell line: You could do the experiment with one cell line, where both drug responsive and resistant strains available. Different passages of the same cell line could be used as replicates. Variability with in cell line is relatively low, therefore fewer replicates required. However the findings could be specific to that particular cell line and results can not be generalizable.
b) Experiment with different cell lines: One can do the experiment with different cell lines, where drug responsive and resistant strains available (paired design). Or alternatively using set of responsive and set of resistant cell lines. In general variability between different cell lines is higher therefore needs more replicates, bur findings are more generalizable.
c) Experiment with patient/animal samples: The variability between patients/animals is much higher than cell lines, therefore needs higher number of replicates. Apart from true biological variability, there is potential to introduce technical variability in sample collections process.

7. How will any findings be validated?

CoIPs/proximity labelling/immunoflourence eg FRET

8. What are possible sources of bias and confounding variables in the experiment?

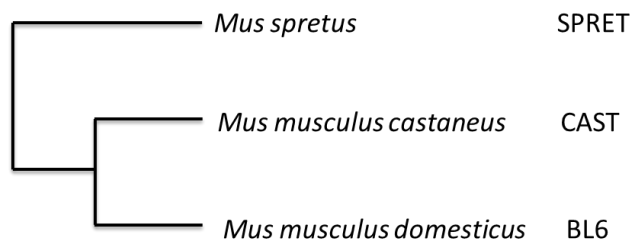Cell line based experiments: Technical sample preparation, Reagents e.g. antibodies, runs etc.
Patient/animal based Experiments: Technical sample preparation, Reagents e.g. antibodies, runs, patient age, sex, sample extraction methods, sample storage methods etc.

9. How can these sources of bias and confounding be controlled?

a) Try to use same procedures and reagents across the experiment.
b) Randomisation
c) Collect and store the metadata associated with samples.

4. ChIP-seq practical: Investigation into the effect of RARα on transcription in breast cancer tissue treated with estrogen.

You are studying the evolution of mouse gene regulation by assaying transcription factor binding. You have stable colonies of three inbred mouse strains housed in the Biological Resource Unit (they are called BL6, CAST and SPRET, see the picture below for evolutionary relationships). These strains are derived from different species and subspecies in the Mus genus. You have access to other data which suggests that transcription factor binding changes between species, and you want to characterise the divergence of transcription factor binding over this short evolutionary distance. You have picked a model tissue (liver), and you plan to use Chromatin immunoprecipitation and sequencing (ChIPSeq) to assay transcription factor binding in all three strains. You have selected four transcription factors to assay plus Polymerase II as a proxy for transcription of genes (HNF1α, HNF4α, HNF6, CEBPα, PolII).

```
┌──────────── Mus spretus            SPRET
┤
│   ┌──────── Mus musculus castaneus  CAST
└───┤
    └──────── Mus musculus domesticus BL6
```

Solutions

1. What is the main aim of the experiment?

   Be precise: To measure differences in transcription factor binding among three mouse strains.

2. What are you measuring?

   Be specific: The number and location of transcription factor binding sites genome-wide.

3. What is your control?

   - The control for a ChIPseq is an aliquot of DNA removed before the chromatin crosslinking step, called an Input. To reduce costs, it is common practice to use one input sample to control multiple transcription factors. However you will need one input per species or cell line to account for differences in the genome.
   - One input sample from each strain (BL6, CAST and SPRET)

4. How will you validate your results?

- It is considered good practice to use a non-NGS based method. Try to validate as many putative targets as practical.
- ChIP-qPCR the on top 15 hits from each strain.

5. What is your study design? How many replicates will you use? Will you use biological or technical replicates? What is the total number of samples you will prepare and sequence?

- You need <u>a minimum</u> of 3 replicates for any statistical analysis. If you start with only 3 replicates and 1 sample fails QC at the final hurdle, you will have to start all over again, what a waste! So always start with more than 3 replicates. As with any project, the more replicates you use the more sensitivity your experiment will have – 6 replicates will allow you to detect smaller differences than 4 replicates.
- At least four mice from each group.
- These are independent mice, therefore they are biological replicates.
- Technical replicates would not add value to this study because you are not interested in assaying technical variability, and they would be as expensive as biological replicates.
- You have three groups (BL6, CAST, SPRET) and five assays (four transcription factors + Pol II)
- There are 63 samples in total: 3 groups x 4 replicates x 5 assays = 60, plus 3 input samples.

6. What are possible confounding factors and sources of bias? The reference genome for mouse was generated from the BL6 mouse strain; does this have any effect on your experiment?

- There are a huge number. How many can you think of?
- BL6 will align to the reference genome better than CAST or SPRET, which could bias any genomic analysis. Try to source genomic variant for the non-reference strains if possible.
- Mouse age
- Mouse sex
- Mouse diet
- Chromatin processing batch
- Sacrifice time of day (due to circadian rhythm)
- Cage
- Mouse handler
- Genome sequence
- Library preparation batch
- Reagent batches
- Litter relationships

7. What information should you collect about your samples to help identify any problems downstream?

8. When will you collect your samples? Which mice will you choose? How will you store the tissue samples? How and when will you perform the Chromatin immunoprecipitation? How and when will you prepare your libraries? Where will you go to find out the information necessary to make these decisions?

9. Where can you go to find out how to test the quality and quantity of your resulting libraries? Who can you ask for feedback as to whether your libraries are likely to sequence well?

10. NGS Libraries are frequently mixed together and sequenced as pools. Your kit only allows you to pool a maximum of 12 libraries. How many pools will you make for your study, and how many samples will be in each pool? Which libraries will you pool together? Where can you go to find out the information necessary to make these decisions?

- Try to use the smallest number of pools possible, and keep the pool sizes as similar as possible.
- You have only 12 indexes therefore the minimum number of pools you can make is 6 (3 pools of 10, 3 pools of 11).
- Make sure sample groups are present in all pools, ask the Bioinformatics core to randomise your samples across pools.
- The Genomics Core!

Advanced

Given that ChIPseq analysis is based on counting reads, and you have decided you need 10 – 20 million reads for your analysis, what type of sequencing will you use, and how many lanes do you need? Where can you go to find out the information necessary to make these decisions?

Note: Our HiSeq 2500 currently provides 150 – 250 million reads/lane

If you were using a kit which allowed you to pool 96 libraries, how many lanes would you need for this experiment? What could you do to reduce the chance of paying for unnecessary sequencing?

- Between 4 lanes and 9 lanes
- You could sequence 1 lane and check the yield. Based on the fact that yield is reproducible between lanes of the same library pool, you can work out how many further lanes you will need and then submit for resequencing.
- ChIPseq is a counting analysis, so longer reads are of no value: HiSeq Single End 50bp
- Since you have 6 pools, you will need an integer multiple of 6 lanes. 1 lane per pool would give you between 13.6 million and 25 million reads per sample, so start with that
- The Genomics Core