THIRTY THIRD CONVOCATION ADDRESS

by

Sir Walter F. Bodmer, *FRCPath, FRS*

Principal, Hertford College, Oxford

and

Head, Cancer & Immunogenetics Laboratory,
Institute of Molecular Medicine, John Radcliffe Hospital,
Imperial Cancer Research Fund, Oxford

2 February 1999

INDIAN STATISTICAL INSTITUTE

CALCUTTA

THIRTY THIRD CONVOCATION ADDRESS
OF THE
INDIAN STATISTICAL INSTITUTE, CALCUTTA
Delivered by
Sir Walter F. Bodmer, *FRCPath, FRS*

Principal, Hertford College, Oxford

and

Head, Cancer & Immunogenetics Laboratory,
Institute of Molecular Medicine, John Radcliffe Hospital,
Imperial Cancer Research Fund, Oxford

2 February 1999

"The Somatic Evolution of Cancer"

Introduction

Many congratulations to all those who are graduating today. This represents an important turning point in your careers, as your training in statistics and its applications is turned into practice at a key stage of your working professional lives. I am most honoured to have been asked to give this convocation address and to follow such a distinguished line of predecessors starting with my own first mentor and academic teacher, R.A. Fisher in

1962. It was through him that I met Professor P.C. Mahalanobis in Cambridge when he visited the Genetics Department in 1961. It is most impressive that the Institute which he started so many years ago, continues with its emphasis on excellence and on the relevance of statistics in its applications. R.A. Fisher's convocation address which he gave shortly after I last saw him when he visited Stanford University in late 1961, and just 6 months or so before he died, emphasised these distinctive features of the Indian Statistical Institute, and in his usual direct manner, commented that "training in statistics ..... is bound to gravitate to an exhibitionism in useless mathematics unless it is linked .... on the one side with fact finding projects....... and on the other side with opportunities to gain first hand familiarity with at least some field in the natural sciences" It was this attitude that led me from an initial interest in statistics, stimulated by the lectures of another previous convocation lecturer, David Cox, to embed my interests in statistics in the particular science of genetics.

Having decided to concentrate on statistics in my third year courses as a mathematics student at Cambridge University, I was led to take an interest in the mathematically and statistically orientated lectures given from the Genetics Department by R.A. Fisher and his colleagues. The summer before starting I was given a reading list of just three books - all by Fisher, namely "The Genetical Theory of Natural Selection", "Statistical Methods for Research Workers" and "The Design of Experiments". That was indeed a challenge and to this day I doubt that I have fully absorbed the significance of all three of these great works. All trained scientists, I believe, should at least understand the basic principles of experimental design as expounded so

2

clearly in R.A. Fisher's The Design of Experiments. And yet how few really do. The modern computer-based statistical packages hardly even leave you having to decide whether a chi-square or a Student's t-test is more appropriate for the data.

As I started my research in genetics, Fisher urged me to get involved in collecting my own data. He was undoubtedly one of the few outstanding statisticians of his generation who actually did generate his own experimental data, largely in the field of genetics. He also urged one to look at the data - a lesson I have never forgotten. Simple visual representations of data can yield so much insight before the application of a battery of complex statistical procedures and tests.

My detailed involvement in cancer research arose because I was asked to direct Britain's major cancer research organisation, a charity, the Imperial Cancer Research Fund or ICRF. Why have a geneticist as a cancer research director? The answer is because cancer is fundamentally a genetic disease at the cellular level.

Historical background

The great German 19th Century pathologist, Rudolph Virchow, was the first to describe cancer as a disease of cells in his famous book of 1858. The cell theory of biological organisms had only comparatively recently been enunciated, and it was Virchow who, for example, coined the term "leukaemia" meaning "white blood" to describe cancer of the white cells of the blood. By the beginning of the First World War in 1914, most of the

fundamental ideas about cancer and its causes, for example, changes in the differentiated state of cells, the role of immune response to the differences between cancer and normal cells, and the potential role of viruses had all been proposed. The most important by far, however, was the notion that cancers are associated with somatic genetic mutations. Therefore genetic mutations in the cells that gave rise to a cancer, that are not present in normal cells which are all derived from the original single cell, namely the fertilised egg. This idea was first clearly proposed in 1916 by Tyzzer and Strong working with transplantable mouse tumours, based on a thesis put forward by Boveri in 1914 that cancer was associated with abnormalities of the chromosomes. At about the same time, it was clearly recognised that individual cancers varied and adapted their growth to the conditions in their host, and this soon led to the fundamental notion that cancer is a somatic evolutionary process.

At the time the ICRF was founded in 1902 there were still many that thought that cancer could be an infection and so there was significant concern when the first Director, Bashford, in initiating a study of the incidence of different sorts of cancers through the world, then represented by the British Empire, organised the sending of samples of cancers to his institute by post. Since that time, a great deal has been learnt about the variations in the incidence of cancer between different countries, and the possible environmental causes of cancer. The overwhelmingly most important cause of cancer world-wide is cigarette smoking, as first demonstrated by its striking association with lung cancer by Bradford-Hill and Doll in the 1950's. Other clear-cut environmental contributors to cancer are exposure to the ultraviolet light in

the sun, the role of certain viruses, for example, particularly in cervical and primary liver cancer, both common in India, and of course dietary factors, though these may be very hard to specify even where the association is very strong, as with cancer of the large bowel, particularly in Western developed societies. Studies of the way in which cancer incidence in migrants soon matches that of the host country rather than the country of origin, add to the overwhelming evidence that the majority of cancers, perhaps 80% or even more, may ultimately be traced to environmental factors. How then can this information be reconciled with the notion that cancer is a genetic disease at the cellular level?

Cancer is essentially, and mostly, a disease of old age. The sharp upward curve of the plot of the incidence of, say, bowel cancer as it changes with age is easily straightened out on a log-log plot, long interpreted by Armitage and Doll and others as indicating a multistep process. The number of steps, based on a simple exponential model, is given approximately by the slope of this age incidence log-log plot, usually about 5 or 6. This gives, at best, a crude measure of the number of rate-limiting steps that may be required for the pathway from a normal cell to a fully developed cancer. What are these steps? The obvious answer is that they are somatic mutations, or possibly also sometimes stable epigenetic changes in gene expression, which are the underlying basis for the somatic evolution of a cancer. The role of environmental factors can then be either to increase the somatic mutation rate, as is clearly the case with the hydrocarbon based carcinogens, for example, found in cigarette smoke, or to promote the growth of mutated

cells, namely augment their selective advantage, as may be the case for many dietary effects.

Cancer genetics at the DNA level

The first clear-cut evidence for a definite genetic step in the development of a cancer was the observation of a specific translocation, namely exchange between two chromosomes, almost universally found in a particular form of leukaemia called chronic mylogenous leukaemia, or CML. The fact that this exceedingly rare event, which can only now be detected with extremely sensitive DNA based techniques at a very low frequency in normal cells, was present in all the cells of the leukaemia provided evidence both for a clonal origin of CML. It starts with the single cell in which the translocation must be presumed to have occurred, followed by the selection that is required for cells with this genetic change to outgrow all others and become the dominant cell population of the leukaemia. These are the hallmarks of a somatic evolutionary process.

The further development of our knowledge concerning the genetics of cancers depended, however, on the application of the revolutionary new DNA technology. It was through this that it became possible to identify the particular DNA sequences, namely genes, involved at the specific translocation point in the leukaemias which were affected by that event and gave rise to its selective advantage. This discovery depended on the identification of a class of genes found in the genome of the oncogenic,

namely cancer causing viruses, first described in 1910 by Peyton Rous. These are genes which, when mutated and brought into a cancer cell by a virus, give the cell an immediate growth advantage. The normal versions of these genes are found in all normal cells and it is when, for example, a mutation switches on a growth factor that stimulates a cell to divide continuously when it should not, that a dominant cancer causing effect is achieved. This may happen not only through the mediation of a virus, but also by a straightforward mutational process. In this way, a number of dominantly acting "oncogenes" have been identified, several of which such as the ras oncogenes, are commonly mutated in specific ways in a variety of human and animal cancers. It is a pair of such oncogenes which is involved in the CML leukaemia associated translocation, of which subsequently many more different examples have been identified and characterised in different types of leukaemias.

Knudson, another geneticist who was a cancer research director, suggested in 1971 the basis for a search for a totally different class of genetic changes in cancers which are fundamentally recessive in their mode of action. He pointed out that, if a cancer arises through a series of somatic genetic changes, then sometimes one of those genetic mutations may be inherited through the germline and so be present in every cell in the body. Such individuals then have all their cells already one step along the somatic evolutionary pathway leading to a cancer, and that head start can be the basis for a dominantly inherited cancer susceptibility. Knudson further suggested that another genetic step was needed somatically, by which the activity of the remaining normal gene was knocked out. Thus, the familial inherited

susceptibility is dominant, while the genetic event that contributes to tumour progression at the somatic level is recessive. These ideas predict a class of recessive genetic changes in tumorigenesis, whose normal functions are to block the development of a cancer and that is why they have sometimes been called tumour suppressors.

Familial adenomatous polyposis (FAP)

Familial adenomatous polyposis (FAP) is a classic example of an inherited susceptibility to a tumour due to mutation in a tumour suppressor gene. FAP was first described late in the 19th century and then identified as a clear-cut dominant Mendelian inherited susceptibility to colorectal cancer in the 1920's. Affected individuals usually develop a few hundred to several thousand pre-cancerous growths or polyps in the colon and rectum, starting usually in their early teens. If left untreated, one or more colorectal carcinomas will inevitably arise in the third or fourth decade. Removal of the large bowel obviates most of this risk if carried out sufficiently early, though it does leave risks associated with certain other manifestations of FAP. It is a classic example of a disease maintained in the population by the balance between mutation producing the abnormality, and natural selection against FAP individuals, a balance first properly analysed more than 70 years ago by J.B.S. Haldane, other famous British geneticist with a long established association with the Indian Statistical Institute.

In the absence of any clue as to the biochemical basis of the disease, the modern approach of positional cloning of the gene led to its identification.

R.A. Fisher, in 1935, and J.B.S. Haldane before him had already in the 1920's and 1930's predicted the value of using linked genetic markers to study disease susceptibility in families. They, however, could not have conceived of the power of this approach that now comes with the DNA-based technology which gives us an essentially unlimited range of DNA variants that can be used to localise a disease susceptibility gene on the human genetic map. This then enables the use of the powerful techniques of genomic analysis to identify the gene that carries the mutations which explain the abnormality as it occurs in families. In this way, a large gene coding for 2843 amino acids, was identified as the culprit that was mutated in FAP individuals. Exactly as Knudson predicted, the mutations giving rise to FAP knocked out the function of the APC gene, namely more than 95% of cases were truncating mutations. Such mutations are found in up to 80% of sporadic bowel cancers, that have nothing to do with any inherited susceptibility. Furthermore, in nearly all the cancers which have these mutations whether in FAP individuals, or sporadic, the remaining wild-type activity is either mutated or lost, so that only abnormal APC genes and their products are found in the cancer.

The distribution of types of APC mutations at the DNA level, and their positions along the gene, provide fundamental evidence about the mutagenic process, environmental factors, and gene function. The spectrum of mutations found in the germline, namely in FAP patients, and in sporadic tumours, namely somatically, is basically very similar and in detail does not show evidence of the effects of environmental mutagens, as is the case, for example, for mutations in other genes in lung cancer. The same is true for

the gene p53, which is very commonly mutated in all sorts of human cancers, including bowel cancers. Its mutations also do not show the effect of environmentally-based mutagens in bowel cancers. This provides strong evidence that the dietary effects so clearly associated with the incidence of bowel cancer are almost certainly not due to mutagenic effects, but are connected with effects on the progression of a tumour, namely the selective advantage of mutations at different stages of a cancer's evolution.

Haldane, in 1935, made the first pioneering attempt to estimate the germline mutation rate of a human gene, using data on the X-linked disease, haemophilia, based on his mutation-selection balance theory. The difficulty at that time, and indeed until recently, was that such a mutation rate was for the whole of a gene and did not distinguish the different rates that may apply to different types of DNA sequences within a gene, or indeed different individual nucleotide positions. We now know, for example, that a C followed by a G is a position usually associated with a relatively high mutation rate, and that the same holds for certain types of repeat sequences where errors are frequently made during DNA replication. Using an extension of mutation-selection balance theory to take into account this sort of variation, and using the data on the frequencies with which different types of germline mutations are found in the APC gene, it has now been possible to obtain much more precise estimates of the mutation rate at the DNA level. It is clear that this can vary by up to a factor of a thousand between the lowest rates for simple types of base substitutions, to an approximately 40-fold higher rate for C's next to G's and a 1000-fold or more higher rate for certain types of repeated sequences that are particularly hard for the DNA

polymerase to copy. These differences are reflected by the spectrum of germline mutations along the gene, which has peaks, for example, at a position where there is a repeat of 5 base pairs. Mutations at this position account for 10-15% of the total of APC germline mutations in a sequence of more than 4000 base pairs. The somatic APC mutations found in sporadic colorectal carcinomas show a somewhat different distribution, with a clustering of mutations in a central region of the gene consisting of only two or three hundred base pairs. Within this are found 50% or more of all the somatic APC mutations. In this case the explanation is most probably that mutations at this position of the gene have a more disruptive functional effect, because of the way that the APC protein interacts with other proteins binding in that middle part of the gene, and thus have a stronger selective effect. This means that mutations in this critical part of the APC gene are more likely to be successfully incorporated into a cancer, following fairly standard ideas derived from conventional population genetics.

The main function of the APC gene appears to be to control the turnover of another molecule, β-catenin, which acts both to control epithelial cell-cell adhesion as well as signalling to the nucleus for growth stimulation. The APC truncation mutations interfere with the normal turnover of β-catenin thus enhancing the growth stimulation signalling pathway and minimising the cell differentiating effects of cell-cell attachment. Thus, the key early role of APC mutations in colorectal cancer development can be accounted for by enabling early independence of growth, releasing cells from the constraint of normal tissue architecture. This derangement may well be the most important early step in many so called solid cancers, such as of the

bowel, breast, stomach and lungs. Subsequent selection then leads to an increased growth rate and provision of nutrients through the development of a blood supply, the process called angiogenesis.

In addition to the obviously disabling APC truncation mutations, which account for the simple Mendelian pattern of inheritance of FAP, there is a small category of missense mutations (in which one amino acid is displaced by another as a result of a simple mutational event) which seem to increase significantly the risk of getting colorectal cancer but without giving rise to the extreme Mendelian FAP phenotype. Such variants show a much lower level of penetrance, namely the risk of getting colorectal cancer associated with them is well under 100% and may be as low as 30-50%. This means that these missense variants are much less selected against in the population than the severe truncating FAP causing APC mutations. From time to time, therefore, their frequency in the population may be much higher than that expected from a mutation-selection balance equilibrium, since they can increase in frequency to appreciable levels just by chance, the phenomenon called genetic drift first analysed by R.A. Fisher in the 1920's. When such mutations occur in the germline they may have a significant probability of contributing to the development of a cancer because they are present in all cells and so a low selective advantage is multiplied up by the number of cells in which the mutation exists. When the same type of mutation occurs somatically in an isolated cells, the probability of it contributing to a cancer is small because of its relatively low selective advantage. This evolutionary perspective can therefore explain why variants such as these, and similar mutations in genes causing susceptibility to breast cancer, can give rise to

inherited susceptibilities which are expressed at the cellular level and yet similar mutations are not found as somatic events in sporadic cancers.

The relatively high frequency of the missense variants means that, in spite of their limited penetrance, they may account for a significant fraction of inherited cancer susceptibility. When coupled with the fact that the severe truncating APC mutations will probably occur as new mutations in 25% or more of FAP patients, there begins to be a significant case for population screening for APC mutations. The technology that would make this possible is being rapidly developed. The case for screening is that benefit can be provided by monitoring at-risk individuals for the development of precancerous growths, which can then be removed when they are found and so obviating the risk of the subsequent development of a bowel cancer. Such population based screening may be expensive and raises the obvious question of the balance between costs and benefits. It also emphasises the need for public education so that people understand why they may be screened and the beneficial consequences of a positive result, since this should prevent them from developing a cancer which might otherwise have only been diagnosed when it was too late to be cured.

The ability to establish an individual's genetically based susceptibility to bowel cancer, or indeed any other disease, raises the important question as to who has a right to know about an individual's genetic status in this respect. Is it indeed only the individual, or is there an obligation to tell a spouse that half their children on average may have a strong susceptibility to bowel cancer, but that this can be avoided with appropriate screening and

minimally invasive procedures. And what about the obligation to tell one's children or other close relatives such as brothers and sisters, or even cousins who may not be aware that they have a significant chance of being at risk genetically of a disease which can be prevented if suitable procedures are followed?

Repair deficiencies and increased mutation rates

There is another major and very different category of inherited susceptibility to bowel cancer which, in contrast to FAP, is often associated in families with the occurrence of many difference sorts of cancers. Also, in contrast to FAP, it is not associated with the presence of large numbers of precancerous growths. A series of elegant studies by workers in Finland, Italy and the United States of America, as well as in Britain, showed, using positional cloning technology, that these sorts of familial susceptibilities were due to mutations in a category of genes that are involved in repairing DNA mismatches which arise due to misreplication of the DNA. Just as in the case of the APC gene, similar mutations are found somatically in sporadic colorectal cancers, though in only 10-20% rather than 80% of sporadic bowel cancers. Following, to some extent, the Knudson tumour suppressor model, second events in these cancers eliminate the remaining normal activity of one of the mismatch repair genes, resulting in a complete inability to repair mismatches which can be detected by a very significantly increased mutation rate in certain types of repeated sequences. This has led some workers to the assumption that it is the high mutation rate that is being selected for, and so supporting a view that high mutation rates are necessary to achieve the accumulation of mutations that is needed for the eventual

expression of a cancer. There is a close analogy here with the proposal by DeVries (who was one of the rediscoverers of Mendelian inheritance at the turn of the century) that evolution could be explained simply by mutation without the need for Darwinian selection. Straightforward theoretical considerations, however, easily show that mutation rate itself cannot be a driving force of evolution, whether at the organismal or somatic level, in the face of natural selection. Mutation rate is a second order phenomenon, since it is the products of cells with an increased mutation rate which will be selected for subsequent to the evolution of the increased mutation rate itself. Can there, therefore, be other direct selective effects of the mutations in the mismatch repair genes that are not connected with an effect on the mutation rate? The answer comes from a consideration of the fact that, as in any population growth, the change in population size depends both on the birth rate, which for cells is their rate of division, and the death rate. Thus, it has come to be realised that "programmed cell death" or "apoptosis" is a very important phenomenon not only in maintaining the balance of normal tissue differentiation, but also in the development of a cancer. There are now known to be mechanisms which monitor, for example, the extent of DNA damage in a cell, and if this is beyond repair lead the cell to commit suicide by the process of apoptosis. Other deviations from the normal state, for example, detachment of a cell from its normal substrate, can similarly trigger the process of apoptosis. It seems probable, thus, that some of the key early steps that allow a cancer to develop through growth independence, may at the same time increase the probability of apoptosis. This therefore imposes strong selection for mutations that can counter this process, and that is the probable explanation of the selective advantages of mutations in the

mismatch repair protein. This has most clearly been shown to be the case for mutations in the p53 gene which, as already mentioned, occur commonly in up to 50% of nearly of all types of human cancers. The main normal role of p53 appears to be to detect DNA damage in the form of breaks in the DNA strands, and to arrest cell division until these breaks have been repaired. If the damage is too extensive, then p53 triggers the process of apoptosis and it appears that the mutations in p53 which are found in cancers, can block or reduce the probability of apoptosis and so this easily accounts for their being selected for. It now seems, incidentally, that much current cancer therapy, either with x-rays or drugs, induces apoptosis and so cell death by suicide rather than by direct killing.

Birth and death models

Most earlier attempts at formulating quantitative models for the cancer process have considered only cell division and not cell death, and have not properly taken into account Darwinian selection. This means that they could not easily account for the long lag periods found, for example, for bowel cancers, often of up to 20 years. These models also could not account for those tumours which are benign and never develop properly to become malignant, namely spread well beyond their initial site of occurrence and metastasise to other parts of the body. A simple model for cancer progression which takes into account cell birth and death rates, as well as the probability of differentiation, shows that mutations affecting either the death or differentiation rates of an intermediate proliferating population of cells often result in a new equilibrium between the number of stem cells, which

are those that keep regenerating a tissue, the number of these intermediate proliferating cells and the number of terminally differentiated cells that are shed into the bowel. In this model, a series of mutational steps can occur which each time increase the number of intermediate cells relative to stem cells, but do not yet lead to exponential growth and therefore frank cancer.

The model first of all can explain the occurrence of a benign phase in the development of a cancer. This phase is associated with the finite step increases in size before exponential growth is achieved. Second, it can account for long lag phases before a cancer arises without the awkwardness of having very low rates of exponential growth, which are extremely hard to envisage over periods of up to 20 years. The epithelial cells lining the gut are replaced, on average about every 2-3 days, meaning that there as many as 100 cell generations per year in this population of cells. This provides extensive opportunity for the production of appropriate mutations with the necessary selective advantage to take a tumour on to its next stage. The lag period is simply explained by the length of time between the successful incorporation of one mutation and that of the next, which may require many cell generations before a mutation with the appropriate selective advantage arises.

Conclusions

Each cancer is an independent somatic evolutionary event involving a series of mutations or epigenetic changes in gene expression, each of which provides a new selective advantage to its progeny cells. The sequence of mutations in any given cancer is a genetic evolutionary pathway. The fact

that certain mutations, such as in the APC gene, tend to occur early whereas others such as in the p53 and the mismatch repair genes, tend to occur later indicates that there are functional pathways within which mutations may have an effect at different stages of the evolution of a cancer. Thus, in those cancers which do not have APC mutations, for example, other genes have been identified in which mutations may have a similar effect.

A recent United Kingdom survey asking the question who were the most significant Britons of the second millennium included Charles Darwin amongst the top half dozen. His fundamental idea of evolution by natural selection had a profound effect on how we view the natural living world and, more particularly, the place of the human species within it. I hope I have been able to convince you that Darwin's ideas when applied at the somatic level of cells within an organism, are just as fundamental for our understanding of cancer, which remains in all countries of the world including India, one of the major scourges of mankind. It is through our improved fundamental understanding of the genetic steps involved in the somatic evolutionary process that gives rise to a cancer, that hope lies in new more effective approaches to its prevention and treatment.

Statistics in one form or another plays a key role in our developing understanding of the fundamental nature of cancer and how we may prevent and treat the disease. This has been a continuing theme during my 20 years in cancer research, which have been most rewarding and stimulating. Once again I congratulate you all on your achievements in statistics, and hope that

you will have as interesting and rewarding an experience in applying your statistical knowledge as I have had.

The following are some key references to my own work in this area:-

Bodmer, W. F., 1996, The somatic evolution of cancer. The Harveian Oration of 1996, *Journal of the Royal College of Physicians of London* **31 (No. 1)**:82-89.

Tomlinson, I. P. M., and Bodmer, W. F., 1995, Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models, *Proc. Natl. Acad. Sci. USA* **92**:11130-11134.

Tomlinson, I. P. M., Novelli, M. R., and Bodmer, W. F., 1996, The mutation rate and cancer, *Proc. Natl. Acad. Sci. USA* **93**:14800-14803.