Use of a Multistrain Assay Could Improve the NTP Carcinogenesis Bioassay

Michael F.W. Festing
MRC Toxicology Unit, University of Leicester, Leicester LE1 9HN, UK

Substantial efforts have been made to develop *in vitro* and short-term tests for predicting mutagenesis and possible carcinogenesis. Such tests already play an important part in the development of new chemicals, though early hopes that they could replace rodent bioassays have not been realized. Ashby and Morrod (*1*) suggested using a more imaginative approach to predict potential human carcinogens, including the use of standard genetic toxicity tests, possible use of the new rodent transgenic mutation assays, and an overall emphasis on mechanistic studies. The latter are of particular importance in assessing human risk from nongenotoxic rodent carcinogens. Genetic variation in the response of rodents to toxic chemicals may have important implications for such a scheme for two reasons. First, all current rodent bioassays use a single strain of mice and/or rats. The extensive Carcinogenic Potency Database on over 450 chemicals, which has been built up over more than 20 years by the National Toxicology Program carcinogenesis bioassay (NTP-CB), mostly using F344 rats and B6C3F$_1$ hybrid mice (*2*), provides the main tool for validating alternatives to the rodent bioassay. At the First World Congress on Animal Testing and Alternatives, held in Baltimore, Maryland, in November 1993, the NTP database was frequently described as the "gold standard" against which all alternatives should be tested. Yet Ashby and Purchase (*3*) found that among 100 chemicals judged to be rodent noncarcinogens, 25 were mutagenic in *Salmonella*, and among 162 chemicals in which there was either clear or some evidence of rodent carcinogenicity, only 91 were mutagenic in *Salmonella*. Such a disappointing correlation between the rodent bioassay and the short-term tests may be due to the inadequacy of the short-term test, but it may also be due to a fault of the rodent bioassay. If some carcinogens have been missed because F344 rats and B6C3F$_1$ mice happen to be resistant to some chemical carcinogens, then the value of the *in vitro* tests may be questioned, as they will appear to be too sensitive.

Second, where strains do differ in response, studies of the reasons for such differences may provide a useful additional tool for mechanistic studies. So far, this approach has rarely been used by toxicologists.

Problems with Current Testing Method

The current NTP carcinogenesis bioassay involves administering the test compound to B6C3F$_1$ hybrid mice and F344 inbred rats for a substantial proportion of their natural life span. However, many forms of human cancer depend both on an environmental insult such as cigarette smoke and on genetic susceptibility. Thus, although smoking is widely regarded as being one of the most important causes of human cancer, most people who smoke do not develop lung cancer. Part of the individual variation is due to genetic factors (*4,5*). In laboratory rats and mice, such genetic variation is most clearly seen as strain differences in response to carcinogens and other toxic chemicals. For example, Shellabarger et al. (*6*) found that when diethylstilbestrol (DES) was administered to Sprague-Dawley (SD) rats, even spontaneous mammary tumors were almost eliminated. Thus, when tested with SD rats, DES would be judged a noncarcinogen. However, in the same study, more than 70% of ACI rats developed mammary tumors when treated with DES, compared with less than 1% in the controls, so when tested in a different strain, DES would be considered a powerful carcinogen. Thus, whether or not a chemical is classified as a carcinogen may depend entirely on the chance of whether the strain chosen for the bioassay happens to be genetically susceptible or resistant to the chemical. The scientific method should reduce and quantify the effects of chance on the experimental outcome. If the outcome of any particular NTP-CB depends largely on unquantified chance, then it can hardly be called a scientific procedure.

Mouse and/or rat strain differences have been reported in response to a wide range of chemicals, including carcinogens such as 9,10-dimethyl-1,2-benzanthracene (DMBA), urethane, 3-methylcholanthrene, 1,2-dimethylhydrazine, 4-nitroquinoline *N*-oxide, 3,4-benzpyrene (*7,8*), teratogens such as cortisone (*9*) and 2,4,5-trichlorophenoxyacetic acid (*10*), and acute toxins.

There are often large strain differences in the response of laboratory animals to toxic chemicals and carcinogens, with some strains being totally resistant to dose levels that cause acute toxicity and/or cancer in other strains. The current National Toxicology Program carcinogenesis bioassay (NTP-CB) uses only a single isogenic strain of mice and rats and may therefore miss some carcinogens. New short-term tests to predict mutagenesis and possible carcinogenesis are validated using data from the NTP-CB. If the animal data are inaccurate, it may hinder this validation. The accuracy of the NTP-CB could be improved by using two or more strains of each species without increasing the total number of animals. It would be possible to continue to use sample sizes of 48–50 animals, but subdivide these into groups of 12 animals of 4 different strains (48 animals total) per dose/sex group, for example, instead of 48 identical animals. This would quadruple the number of genotypes without any substantial increase in cost. Such a multistrain "factorial" design would, on average, be statistically more powerful then the present design and should increase the chance of detecting carcinogens that currently may give equivocal results or go undetected because the test animal strains happen to be specifically resistant. When strains differ in response, studies of differences in metabolism, pharmacokinetics, DNA damage/ repair, cellular responses, and in some cases identification of genetic loci governing sensitivity may provide biological information on toxic mechanisms that would help in assessing human risk and setting permissible exposure limits. The NTP may have made the world a safer place for F344 rats and B6C3F$_1$ mice. The challenge now is to show that the results can be generalized to other genotypes and that the database can be used to validate short-term tests. A first step could be to explore the importance of strain differences in the carcinogenesis bioassay. *Key words*: carcinogenesis testing, National Toxicology Program, rodent bioassays, strain differences. *Environ Health Perspect* 103:44–52 (1995)

For example, Pohjanvirta et al. (*11*) found that the LD_{50} of TCDD varied by two orders of magnitude between two strains of rats. A strain that is highly sensitive to the acute toxic effects of a chemical may have such a low maximum tolerated dose that no extra tumors are observed during the normal carcinogenesis screening period. Table 1 gives some additional examples of strain differences in response to carcinogens. For example, Lijinsky et al. (*12*) applied methylnitrosourea to the shaved skin of groups of 20 mice from 4 strains. In strain CD-1 only three mice developed tumors (15%), whereas all the BALB/c mice developed them. The statistical implications of these results are discussed later.

In the past such strain differences have been of great importance in stimulating research into toxic mechanisms. Much of the work on cytochrome P450 and the *Ah* locus followed observations of large mouse-strain differences in response to carcinogens such as 3-methylcholanthrene associated with the inducibility of arylhydrocarbon hydroxylase (*13*). Similarly, over a period of some years, Drinkwater and colleagues (*14*) have been studying the mechanisms of hepatocarcinogenesis in mice, based on the finding that strain C3H is susceptible and C57BL/6 is resistant to the induction of liver tumors by a variety of carcinogens. Much of this difference seems to be due to a single genetic locus designated *Hcs* (for hepatocarcinogen sensitive), which may affect the proliferative rate of both normal and preneoplastic hepatocytes (*15*). These two examples clearly show that strain differences can be used in the study of toxic mechanisms, even though a full understanding usually requires further extensive research.

In view of such large strain differences in response to known powerful carcinogens, the wisdom of using a single strain of mice and/or rats is questionable. Weindruch and Masoro (*16*) expressed grave reservations about the overuse of F344 rats in gerontological research, pointing out that the extent to which results obtained with a single strain can be generalized is unknown. They went on to suggest that "It seems appropriate to encourage the development of several genotypes in ageing research. . . ." (*16*: p. 88). Yet carcinogenesis screening, which is of much more immediate importance to human health, continues to be done using a single strain of mice and rats, and those involved with the NTP-CB appear to be making no effort to devise a better strategy.

Optimum Deployment of Resources

In the current NTP-CB the test chemical is administered to groups of 50 mice or rats of each sex. Each group consists of a

Table 1. Some examples of strain differences in response to carcinogens

Carcinogen	Comment	Reference
Methylnitrosourea	Applied to shaved skin of groups of 20 female mice treated at 8-weeks old. Tumors were 40% in Swiss (10% in Swiss controls), 70% in SENCAR (5% in SENCAR controls), 15% in CD-1 (0% in CD-1 controls), and 100% in BALB/c (0% in BALB/c controls).	(*12*)
7,12-dimethylbenz[*a*]-anthracene	Gastric intubation in 10 rat strains. No mammary tumors seen in the COP strain, 1.2 tumors/rat in the F344 strain, and 5.0 tumors/rat in the OM strain.	(*54*)
3,2'-dimethyl-4-aminobiphenyl	50 mg/kg given subcutaneously for 10 weeks. Tumors of prostate seen in 48% of F344, 41% of ACI, 13% of LEW, 7% of CD, and 0% of Wistar rats.	(*55*)
N-methyl-*N*'-nitro-*N*-nitrosoguanidine	Single oral dose given to 5-day-old or 28-day-old ACI or WF strain rats. 58% tumors in glandular stomach seen in ACI rats, none in WF. Forestomach tumors seen in both strains. Lung tumors in 64% of WF rats treated at 5 days, but none in ACI rats.	(*56*)
N-methyl-*N*'-nitro-*N*-nitrosoguanidine	100 µg/ml in drinking water for 30 weeks. Adenocarcinomas of glandular stomach in 60% of SD, 67% of WKY, 53% of LEW, 23% of Wistar and 6% of F344 rats at week 50.	(*45*)
Diethylnitrosamine	Single intraperitoneal dose at 8 weeks of age. 50% heptaocellular carcinomas seen in strain R16 rats but only 3% in strain ACP rats.	(*57*)
Iron plus hexachlorobenzene	90% hepatocellular carcinoma seen in C57BL/10 mice, but none in DBA/2, though the experiment in the latter strain had to be terminated after 10 months due to excessive mortality.	(*48*)

control, a top dose level given the maximum tolerated dose (MTD; defined as the highest dose of the test agent given during the chronic study that does not alter the animals' normal longevity from effects other than carcinogenicity) and two more groups which usually receive 50% and 25% of the MTD, though this protocol has varied slightly (*17*). Each assay uses 400 animals of each species and costs approximately $1–2 million or more, though this can vary widely depending on dosing methods and other factors. It would be unacceptable for an "improved" assay to use any more animals or to cost substantially more. Thus, if it is accepted that there may be large strain differences in response, the question is, what arrangement of genotypes within each dose group of 50 animals would be optimum for detecting a carcinogen, given that the response may be under genetic control?

Three Alternatives

Outbred stocks. "Outbred stocks" of laboratory mice and rats are closed colonies which are usually propagated by some form of random or haphazard mating, or sometimes by a breeding system that avoids the mating of close relatives. The effect is a colony of animals that is reasonably uniform in its characteristics, though each individual is genetically distinct. Outbred stocks have sometimes been compared with the genetically variable human population, but this is misleading because

often the amount of genetic variation is substantially less than that found in humans. Outbred stocks are often known by generic names such as "Wistar" and "Sprague-Dawley" rats and "Swiss" mice, or alternatively by a more specific code name given by the breeder. However, as a result of genetic assortment, inbreeding, and selection, different colonies of Wistar and Sprague-Dawley rats and Swiss mice will be genetically different from one another.

Many toxicologists have favored the use of genetically variable outbred stocks because they contain more than one genotype (*18*). For example, Arcos et al. (*19*) considered that it is more correct to test on a random-bred stock on the grounds that it is more likely that at least a few individuals will respond to the administration of an active agent in a group which is genetically heterogeneous. There are a number of serious disadvantages in using outbred animals for carcinogenesis screening, but this does not invalidate the desirability of using more than one genotype.

One problem is that the actual degree of genetic variability depends on the past history of the strain and in many cases is quite limited. For example, Festing (*20*) used DNA fingerprinting to study genetic variation among 10 inbred strains and 6 outbred stocks of rats, including 3 colonies of Wistar, 2 of Sprague-Dawley and 1 Hooded Lister stock. Genetic variation was assessed by the degree of band sharing

between any two individuals (*21*). The members of an inbred strain are like identical twins and, as expected, there was little variation within inbred strains (approximately 100% band sharing), except when colonies had been kept separately for several years or there was genetic contamination. The degree of band sharing across the 10 different inbred strains was 34 ± 15%, which is about what would be expected among humans picked at random. In contrast, within the six outbred stocks the band sharing ranged from 84 to 95 ± 5%. This implies that within these stocks there was a high degree of genetic uniformity. In all samples there were pairs of rats that had identical fingerprints (Fig. 1), a result that would be virtually impossible in a small sample of humans (assuming identical twins were excluded). Thus, the level of genetic variability within outbred rat stocks falls far short of that found in human populations, whereas differences between different inbred strains seem to be of the same order of magnitude as differences between individual humans. Few studies have been conducted on genetic variability in outbred mouse populations. The most detailed study was that of Rice and O'Brien (*22*), who found that there was still substantial genetic variation in three outbred mouse colonies, even though they had been maintained as closed colonies for several years. Whether these conflicting findings reflect a true species difference or are a reflection of the different methods of quantifying genetic variation remains to be determined. It is possible that rat colonies have, in the past, been maintained with fewer breeding pairs, leading to closer inbreeding, simply because rats are larger than mice and take up more breeding space.

In any case, it is technically simple to produce a genetically variable stock by crossing several strains. However, a stock with uncontrolled genetic variation pro-
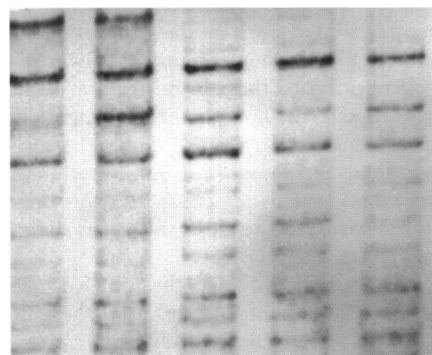

Figure 1. DNA fingerprint of seven outbred hooded Lister rats. Note that 5 of the rats have about 14 bands that can be resolved, and all bands are identical. Two other rats have a slightly different pattern. The chance of two human DNA fingerprints matching in this way is very low, provided identical twins can be excluded.

duced in this way is not ideal because it is impossible to ensure that treated and control groups are genetically identical at the start of the study. All carcinogenesis bioassays are done against a background of spontaneous tumors, many of which are genetically determined. Using an outbred stock may occasionally result in more genetically determined spontaneous tumors occurring in the treated than in the control groups by chance. These would be mistaken for chemically induced tumors, leading to a false positive result (i.e., the chemical is classified as a carcinogen when it is not one). Also, unless the pedigree of each individual animal is known, which is usually impractical, it is impossible to know whether any variation in response to the test chemical is due to genetic factors. There are several other disadvantages in using outbred stocks. Over a period of a few years the genetic composition of colonies is likely to change due to selection and genetic drift, so the characteristics of the stock will change. This can be minimized by maintaining large colonies with breeding schemes designed to minimize inbreeding and genetic drift, but such schemes can be difficult to administer and can be expensive. Genetic quality control of outbred stocks, aimed at ensuring that they have not become mixed with other stocks, is difficult. Consequently, few outbred colonies are monitored. At present there are not even any genetic markers that will distinguish between Wistar and Sprague-Dawley rats. These have to be taken on trust from the supplier.

Isogenic (inbred and F₁ hybrid) strains. Inbred strains have been developed by many generations of brother × sister mating with all individuals tracing back to a common ancestral breeding pair in the 20th or a subsequent generation (in order to eliminate the many parallel branches that could develop). The resulting inbred strain is almost like an immortal clone of genetically identical individuals which are also homozygous at virtually all genetic loci. Genetic variability within the strain is eliminated, leading to greater phenotypic uniformity, so statistical precision is increased. The strain stays genetically constant for long periods of time so that background data collected on such strains remains valid for longer periods, provided that the environmental conditions remain constant. Genetic quality control methods to ensure that the correct strain is being used are much easier to apply than in outbred stocks. Each inbred strain is designated by a code consisting of uppercase letters, and sometimes numbers, such as C57BL and F344, following rules developed by the Committee on Genetic Nomenclature of the Mouse (*23*) and a similar committee for the rat (*24*).

F_1 hybrids are the first generation cross between two inbred strains. Like inbred strains, they are isogenic (i.e., all individuals are genetically identical), and have all of the useful properties of inbred strains except that they are not homozygous at all genetic loci, and they tend to be more vigorous than inbred strains.

The NTP-CB uses two isogenic strains for carcinogenesis bioassays: the B6C3F$_1$ hybrid mouse and the F344 inbred rat strain. However, the use of a single isogenic strain is rather like using a clone of genetically identical humans in a clinical trial. Given that humans vary genetically, no single individual could represent the human race. If the isogenic strain happens to be relatively resistant to the carcinogenic effect and/or highly sensitive to acute toxicity by the chemical, leading to a low MTD, then no increase in tumor incidence may be seen by the end of the bioassay, giving a false-negative result (i.e., a rodent carcinogen is classified as a noncarcinogen).

Multistrain experiments. A third possibility, which appears to offer many advantages (*25,26*), is to use several strains, but without increasing the total size of the experiment. In effect this is a bit like creating a "synthetic" outbred stock with many of the advantages of an outbred stock in terms of genetic variability, but with few of the disadvantages. The strains would be mixed in defined groups, without losing the identity of each individual. For example, instead of using 50 animals per group, all of the same isogenic strain, it would be possible to use 48 animals per group, but with a mixture of, say, 12 animals of each of four different isogenic strains individually randomized to each treatment group. This would immediately quadruple the number of genotypes and reduce the chance that all of the animals were resistant to the test compound. The introduced between-strain variation would usually be much greater than the within-strain variation found in outbred stocks (*27*: p. 241). This design would avoid one of the main problems associated with the use of outbred stocks, as the treated and control groups would always be genetically identical at the start of the experiment, thereby reducing the chance of false-positive results. This scheme would have all the advantages of using isogenic strains, such as long-term genetic stability and phenotypic uniformity, but without the major disadvantage of using animals of only a single genotype. It should not significantly increase the cost, as the total size of the experiment would be the same, and theoretical studies show that statistically it would normally be more powerful than using a single isogenic strain (*28*). It would

also show whether the response is highly polymorphic or is uniform across strains. Tennant (29) has suggested that compounds that cause cancer in animals of many genotypes are more likely to be human carcinogens than those which only cause cancer in limited range of genotypes. The multistrain bioassay should give a much better estimate of the extent of genetic variation in response.

There may be other substantial advantages. For example, strains that are relatively susceptible to a carcinogen usually have a shorter latency to develop tumors. In some cases, it may be possible to terminate a study earlier, thereby saving time and money. For example, Smith and colleagues at the Medical Research Council Toxicology Unit in Leicester, UK, have found that the latency to develop liver carcinoma after dietary administration of tamoxifen (a known rat liver carcinogen) was approximately 20 months in F344 rats (the strain used in the NTP-CB), whereas there was a 100% incidence in LEW and Wistar rats after only about 11 months. A saving of 9 months in obtaining a positive result may be of considerable interest both to the NTP-CB and to the pharmaceutical industry.

More than 25 years ago, Heston clearly stated the case against outbred stocks and for using more than one isogenic strain:

> Yet the question is sometimes asked, why not use genetically heterogeneous stock mice so the results will be more applicable to the genetically heterogeneous human population? The answer is that we are not trying to set up a model with mice exactly comparable with humans. This would be impossible because mice and men are different animals. What we are trying to do is to establish certain facts with experimental animals and this can be done, or done more easily, when the genetic factors are controlled. Once the facts are established we then, with much common sense, see how the facts can be related to man.
>
> When genetic variability is desired this can be obtained in the highest degree by using animals of a number of inbred strains. This variation between strains is usually much greater than is found in animals of a non-inbred stock which actually may be rather uniform although more variable than an inbred strain. (30: p. 5)

These words are still true more than a quarter of a century after they were first published.

An Additional Method for Studying Toxic Mechanisms

A potentially important benefit of the multistrain carcinogenesis assay is that it may provide an additional way of studying toxic mechanisms. In the past, animals used in the rodent bioassay were virtually regarded as a "black box." If they developed tumors,

then the test substance was regarded as a (human) carcinogen, and appropriate safety precautions were specified. If no extra tumors were observed, the substance was regarded as a noncarcinogen. Recently, however, much more emphasis has been placed on understanding the underlying mechanisms (1). The data are interpreted in the light of detailed pharmacological and toxicological studies aimed at a much fuller understanding of the fate of the chemical in the body. It is even widely accepted that some rodent carcinogens, such as certain hydrocarbons causing kidney tumors in male rats, associated with the male rat urinary protein α_{2u} globulin, may not be carcinogenic in humans. The NBR strain of rats, which has a genetic defect preventing it from producing α_{2u} globulin, is resistant to the induction of these tumors (31). This is a simple example of the way strain differences in sensitivity may be used to explore mechanisms.

The use of strain differences to study mechanisms has been applied successfully as a research tool, but it has rarely been used in screening, possibly because it is rare to use more than one strain. The studies of liver carcinogenesis caused by tamoxifen, mentioned above, illustrate the way in which strain differences can provide additional information that may be of value in understanding mechanisms.

Tamoxifen is used to treat women with breast cancer and is now undergoing clinical trials as a chemopreventive agent for women at high risk of developing breast cancer. However, it is known to cause liver carcinoma in rats. Thus, it is important to gain a clear understanding of mode of action and whether this mechanism is likely to be operative in women.

Tamoxifen was administered in the diet (420 ppm, corresponding to approximately 40–50 mg/kg/day in F344 rats) to F344, Wistar, and LEW strain rats (32). Fifteen treated and control rats (5 per strain) were sacrificed at 3 and 6 months for detailed pathological and biochemical investigation, and the rest (10 rats per strain in treated and control groups) were kept until they showed evidence of tumors or loss of condition. By 10 months, all the tamoxifen-treated Wistar and LEW rats had developed liver carcinoma, but none of the F344 rats developed evidence of tumors until at least 20 months of treatment, and these tumors appeared to be less aggressive, involving a smaller proportion of the liver. Thus, the F344 rats had an increase in tumor latency of about 10 months. Had tamoxifen been a weaker carcinogen, or had F344 had a slightly longer latency, tamoxifen may have been classified as noncarcinogenic in this strain.

The longer latency to develop tumors in F344 rats (32) was associated with a

reduced tamoxifen intake, increased blood levels, lack of development of multidrug resistance (33), a delayed build-up of DNA adducts, and a reduction in cell division in the liver when compared with the other two strains. At present, it is not clear which, if any, of these differences are of critical importance. Although the use of more than one strain has not solved the problem of the exact mode of action of tamoxifen in causing liver cancer in rats, it has provided a framework for further investigation, and it did not increase the total size of the experiment: 15–20 control and treated rats at each time point would have been used in any case. It also shows that there is genetic variation in the response to tamoxifen in rats which was not apparent when a single strain was used. If such variation is present in rats, then it may also be present in humans. With current rapid advances in molecular genetics, it may be possible to identify some of the genes involved, and their human homologues could be identified. Although this would be a major research project, the identification of genes controlling susceptibility to carcinogens in laboratory animals and the identification of human homologues appears to be a promising long-term research goal.

Tamoxifen is not known to cause liver cancer in mice. However, this may be because most mice die by about a year of treatment due to overgrowth of spinal bone causing kyphosis. In a multistrain trial, all C57BL/6 and DBA/2 mice died in this way before they were a year old. However, B6C3F$_1$ hybrid mice were much more resistant, and continue to survive tamoxifen treatment beyond 22 months of age. Whether they will develop tumors is not yet known, but at least it should prove to be possible to treat this strain of mice for a large proportion of their normal life span. In this case, the multistrain experiment is better because of the discovery of a strain that is resistant to the noncarcinogenic toxic effects of the compound, making it possible to carry out a reasonable carcinogenesis bioassay.

More research is needed to develop these methods. However, with recent rapid advances in genetics, the use of strain differences to study mechanisms appears to be ripe for exploitation.

Examples of Multistrain Studies

The theory of multistrain toxicity screening has been discussed in a number of publications (18,25,30,34–36) and at several conferences, and a number of such studies have been done. For example, Heston et al. (37) studied the effects of the antifertility drug Enovid in five mouse strains and found no evidence that it was carcinogenic.

More recently, Wolff et al. (*38*) studied the response of four F_1 hybrids to 2-acetyl-aminofluorine using a design of the sort proposed here. They found only small differences between the hybrids, possibly because F_1 hybrids tend to be intermediate in phenotype. They confirmed theoretical studies that the statistical precision was similar using 48 B6C3F$_1$ mice or 12 mice of each of the 4 hybrids, as would be expected when there are no important strain differences. Kalter (*9*) used several strains in a study of susceptibility to the induction of cleft palate by cortisone acetate and found that C57BL/6 was 9- to 34-fold more resistant than strain A/J.

Holson et al. (*10*) used four inbred strains and one outbred stock in a multi-replicated dose–response study of the teratogenic effects of 2,4,5-trichlorophenoxy-acetic acid (2,4,5-T). They found that C3H/He and BALB/c mice were 3.5- to 5-fold more resistant to 2,4,5-T than the A/J strain, though the dose–response lines among all strains were parallel, and they concluded that similar mechanisms were involved with all strains. They did not think that the strain differences were great enough to outweigh the practical difficulties of using several strains in a teratology study, but it could not have been foreseen that a teratogenic effect would be observed in all strains. Moreover, following Tennant (*29*), the fact that a teratogenic effect was observed in all genotypes would suggest that 2,4,5-T is more likely to be a human teratogen than if the effect was only observed in a single strain.

Consequences of the Multistrain Experiment

Although a reasoned scientific discussion of possible disadvantages of multistrain experiments has not been published, a number of potential objections have been discussed at scientific conferences.

More chemicals would be judged to be carcinogenic. If the rodent carcinogenesis bioassay is improved, more chemicals will be judged to be rodent carcinogens. There is real concern that too many chemicals are already classified as rodent, and therefore presumed human, carcinogens (*3,39*). This is sometimes used as an argument against improving current testing methods. While this is a genuine worry that needs to be taken into account in introducing better methods, it should not be used as an excuse for inaction. In the real world, humans are continually being exposed to natural and man-made carcinogens. In many cases this is of little significance because the dose levels are very low and/or the carcinogens are relatively weak. Eventually, the carcinogenic potential of a xenobiotic may have to be assessed on a

quantitative or semi-quantitative scale of potency rather than the present dichotomous scale of "carcinogen/noncarcinogen." In practice, this already happens with many compounds that are known rodent carcinogens still being widely distributed, with suitable precautions to ensure reasonable safety standards.

It should also be recognized that risk assessment is a two-step process. The carcinogenic potency or hazard of a chemical must first be established in animals that have been chosen because observations in these animals are believed to be relevant to humans. The animal data are then used to assess human risk, taking into account all available biological information (*30*), which may even suggest that the mechanism by which the chemical causes tumors in rodents is not relevant to humans. However, if the animal data are inaccurate, they may well lead to contradictory conclusions. For example, the *in vitro* and short-term tests may predict that the chemical will be a genotoxic carcinogen. If the rodent bioassay then gives no evidence of carcinogenicity, this may be because the rodent bioassay is inaccurate, or there may be real problems with the short-term tests which need to be explored. Unfortunately, the rodent bioassay is so expensive that it is unlikely to be repeated, and there is a natural tendency to assume that the animal rather than the *in vitro* test provides the definitive result.

Practical difficulties. Potential practical problems arise from the increased complexity, the danger of strains becoming confused, and the extra cost of multistrain tests. However, strains could be chosen that differ in coat color, or some other easily detected genetic marker, and the animals can be physically marked, so the chance of getting them confused would be minimal. There would be no necessity to start the assay on all the strains simultaneously so, assuming four strains were to be used, the experiment could be divided up into four smaller experiments, which would help to spread the workload. Objections on the grounds of the cost are sometimes raised. The cost of the NTP-CB is largely due to the cost of maintaining the animals and the subsequent costs associated with the pathological studies. These costs are largely dependent on the total number of animals, which would not be altered. However, metabolic studies would need to be done on all strains, which may increase the cost, though again in most cases the total number of animals should not need to be increased. In the end, any extra costs must be equated with the advantages of such a design.

Scientific questions. Scientific questions that need further discussion include the number of and choice of strains and the

problem of strain differences in the MTD.

The optimum number of strains will only be determined in the light of practical experience on the real extent of strain differences. Two strains would be better than one, but more strains would be needed to explore relationships between variables such as rate of metabolism and sensitivity. Four strains might be a useful starting point, with numbers modified by later experience. Rao et al. (*40*: p. 390) favored the use of an F_1 hybrid on the grounds that "... it has a major advantage in that the level of heterozygosity may more closely resemble that of the noninbred individual of the species." However, there is no evidence that the level of heterozygosity has any bearing on response to carcinogens. F_1 hybrids tend to be intermediate in response, which may be an advantage if a single strain is used, but if several strains are to be used, and the aim is to get as wide a range of different genotypes/phenotypes as possible, then they should not all be F_1 hybrids.

Theoretically, different strains might be chosen for each study on the basis of known susceptibility to compounds chemically related to the compound being tested. This approach was used by Heston et al. (*37*) in their study of Enovid. They gave detailed reasons for choosing each of the five strains. Another alternative, suggested by Wolff et al. (*38*), would be to use a subchronic assay to pick the most appropriate strains. No correlation has been established between acute toxicity and carcinogenesis, but whether or not there is some association between the incidence of preneoplastic foci and susceptibility in different strains, as found in studies with tamoxifen, remains to be evaluated.

A good case could be made for using a fixed panel (as is current practice) of, say, four rat and mouse strains chosen because they have a long life span, differ genetically, and have no serious defects. A suitable panel of mouse strains might be B6C3F$_1$ (for continuity), C57BL/6, and DBA/2, two strains favored by the National Institute on Aging (*41*), and either B6D2F$_1$ or BALB/c. Suitable rat strains could include F344 (for continuity), BN, F344BNF$_1$ [all three of which are favored by the NIA (*41*)] and possibly WAG. None of these strains (including the B6C3F$_1$ mice and F344 rats) are entirely free of disadvantages, and other candidates could also be considered. Should genetically engineered strains with, for example, human drug-metabolizing enzymes become available in the future it would be possible to substitute them for one of the strains. In this way a multistrain design could evolve as new animal models become available, while retaining ties with the past

in the form of the F344 rat and the B6C3F$_1$ mouse strains.

Another problem is that the MTD would probably be strain dependent. Multistrain subchronic studies would be needed to determine the MTD for each strain and, either each strain would have to have a different dose level, or they could all be given the dose of the most sensitive strain. The latter approach would probably be quite acceptable unless the MTDs differ markedly between the strains. In effect, this is what happens in the pharmaceutical industry, which largely uses genetically variable outbred stock. In this case the MTD would tend to be biased in favor of the most sensitive individuals. Further research is needed on this topic.

Statistical validity and the problem of group size. Toxicologists frequently express unease about the statistical validity of a multistrain design. Part of the problem seems to be differences in perception of the size of the groups that are to be compared.

Many toxicologists immediately look for a group of animals that are identical in every respect except for the applied treatment and regard this as the basic unit for comparing treatments. Thus, with a single isogenic strain and 48 animals per group, they would be comparing 48 control animals with 48 animals in one of the treatment groups. However, with a 4-strain experiment they assume that they would only be comparing 12 control animals of each strain with 12 animals in the treated group, and this would have to be done four times. Inevitably, they consider group sizes of 12 animals as being too small, so they reject the multistrain experiment on the grounds that it lacks statistical power. This is a misunderstanding of the concept of what constitutes a "group." The problem can be clarified by considering the following: Assume for simplicity that a screening experiment involves only a treated and a control group, with 48 animals in each. These groups can be set up in various ways:

- Single isogenic strain: If a single isogenic strain is used, with 48 animals per group, then it is clear that the group size is 48 animals.
- Single nonisogenic strain: What is the group size if a single nonisogenic (outbred) stock is used? In an outbred stock all individual animals are genetically different. Does this mean that if an outbred stock is used there are 96 groups of one animal? Clearly not. Everyone who uses outbred animals still considers that they have two groups of 48 animals, so clearly a "group" does not have to be genetically homogeneous.
- Single nonisogenic strain, animals typed at a marker locus: Suppose next that

somebody types the outbred animals for a particular genetic marker after an experiment is concluded and finds that half the animals of each group are type A and the other half are type B. There are now two treated and two control groups of 24 animals instead of two groups of 48. Would this experiment now be less powerful just because somebody had typed the animals genetically? Could it be statistically more powerful not to type the animals? Can ignorance be more statistically powerful than knowledge? Or can it be argued that it makes little difference whether the animals have been genetically typed, the comparison between the treated and control groups is still based on 48 animals per group? A little reflection suggests that the latter argument is valid, but that if the animals have been typed, then this adds some information which may be useful. The basic comparison would still be between a treated and a control group of 48 animals, but if it was found that tumors were only observed in treated animals of type A, then this would provide additional information that was not available before the animals were typed. Clearly, in some circumstances it is biologically more powerful, and statistically no less powerful to know the genotype of the animals.

- Multistrain, "blind" experiment with 4 strains done: Suppose next that a multistrain screen is set up with 48 animals in the treated and control groups, consisting of 4 strains with 12 animals per strain, but that the experimenter is "blind" with respect to strain of animal. Suppose the experimenter observes 0/48 tumors in the controls and 12/48 in the treated group (or alternatively 0/48 in both groups). Is this experiment invalid as a carcinogenesis screen? The answer is no. There does not seem to be any reason why the treated and control groups should not be compared in this way. It has been established above with nonisogenic stocks that the groups being compared do not need to be genetically isogenic, and given that the experimenter is blind with respect to genotype (just like people who use nonisogenic stocks), he or she has little alternative to comparing the 48 treated with the 48 control animals. This experiment is really no different from one done with outbred animals. However, this would not be a sensible design in practice because valuable biological information on the genetic identity of each individual would be wasted.
- Multistrain, "unblinded" experiment: Suppose now that the strain identity in the above experiment is decoded. Does this extra information now mean that the experiment becomes less powerful? Must

the experiment now be analyzed as four separate experiments, each of which is too small, or can it still be regarded as 2 groups of 48 animals, but with the extra information on the relationship between the individual animals? This extra information may show, for example, that all of the extra tumors occurred in one particular strain, or it may show that the extra tumors were evenly distributed among all four strains. In either case this extra information clearly has biological significance. It would surely be strange if extra information of this sort made the experiment less powerful. And, of course, it does not do so. The experiment can still be analyzed as a comparison of 2 groups of 48, but with additional statistical tests of whether all strains respond in the same way or whether there is evidence that the response is strain dependent. A full discussion of the exact methods of statistical analysis of such multistrain experiments is beyond the scope of this paper. Felton and Gaylor (*28*) suggested that one method of analysis would be to use the Mantel-Haenszel test as a means of combining the data from different strains. This is discussed briefly below. However, it should be clear from the above discussion that the main comparison of interest is whether the number of tumors in the 48 treated animals is greater than in the 48 controls, and the fact that each group is composed of 4 strains is only of secondary importance.

But the multistrain experiment could actually be more powerful. On average, the multistrain experiment should be more powerful than the single-strain experiment. That is one of the reasons why it should be investigated for the NTP-CB assay.

The importance of sensitivity. The above arguments show that genetic heterogeneity as such does not automatically lead to reduced statistical power, though strictly this is only true if the genotypes are exactly balanced between the treated and control groups. This is possible with a multistrain experiment but not possible when an outbred stock is used, as the genotype of each outbred animal is unknown. Thus, the genetic heterogeneity of outbred stocks reduces statistical power, which is the main justification for the use of isogenic strains in the NTP-CB.

However, on average, the multistrain experiment will be more powerful than the single-strain design. The reason for this is that the multistrain experiment increases the chance of including some genetically sensitive individuals in the test population. Felton and Gaylor (*28*) did a detailed study of the statistical power of multistrain compared with single-strain protocols using the Mantel-Haenszel test. A single-

strain protocol results in a 2 × 2 table of tumors/no-tumors in the treated and control groups for each dose-by-sex subgroup. A four-strain protocol would result in four such tables, which need to be combined to obtain an overall estimate of statistical significance. The Mantel-Haenszel test (42) provides a method of combining such tables. Felton and Gaylor (28) did computer simulations involving 1000 samples with 48 control and 48 treatment observations in each case with a range of background tumor incidences and response rates. They compared the single-strain protocol with a 2-strain (24 animals per strain), a 4-strain (12 animals per strain), and a 24-strain (2 animals per strain) protocol using a 1-sided Mantel-Haenszel test with a significance level of $\alpha = 0.05$. The overall conclusions were that

> For the case where there is no knowledge of the sensitivities of available strains, the best design, in terms of the Mantel-Haenszel test, will generally consist of using as many strains as possible. The risk of using such a design is the possible loss of a small amount of power where the average increase in the response rate due to a chemical is small and the power is small anyway. An advantage of the multistrain experiments is the possibility of a large gain in power. (28: 409)

The multistrain protocol was slightly worse than the single-strain one with weak carcinogens and when all strains were identical in response (e.g., causing an increase in tumors of about 5% in all strains). However, in these circumstances the chances of any protocol detecting the carcinogen would be low. The multistrain protocol was of about equal power when the response rate was higher and the strains did not differ but was substantially more powerful than the single-strain protocol when strains differed. Thus, if the examples given in Table 1 are representative of the true biological situation, the multistrain protocol would normally be much more powerful than the single-strain one. The reason for this is that statistical power does not only depend on sample size. The sensitivity of the test animals is also of critical importance. If the animals are resistant to the induction of tumors with a particular agent, as SD rats were to DES when tested by Shellabarger et al. (6), then the substance will go undetected as a carcinogen however large the sample size. However, only small numbers of highly sensitive animals are needed to give a statistically significant increase in the number of tumors.

Table 2 shows the sample size that is needed to give a 90% chance of detecting a difference in tumour incidence between a treated and control group and declaring it statistically significant at the 5% level of probability (one-tailed test) for two levels of spontaneous tumors in the controls and for various levels of response in the treated group [calculated using a program by Piantadosi (43), using methods somewhat similar to those described by Snedecor and Cochran (44), but based on an algorithm for an exact method for calculation of binomial confidence limits]. Note that the sample size needed to detect an effect declines dramatically as the proportion of affected animals increases. With a response of 10% in the treated group against a background of 1% in controls, the required sample size is 108 animals in each group, whereas with a 50% incidence in the treated group the sample size needed is only 12 animals. This can be related to the data of Tatematsu et al. (45), who administered *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) in the drinking water (Table 1) and found a 60% incidence of tumors in SD rats, but only a 6% incidence in F344 rats. Assuming that these incidences represent the true response rates in these two strains and that there is a 1% incidence in controls, sample sizes of 229 F344 rats in both control and treated groups would be required to give a 90% probability of detecting a 6% incidence of tumors, whereas a sample size of only 9 SD rats in each group would be needed (calculations based on 43). Clearly, there would be a high chance that the NTP-CB assay using F344 rats would miss MNNG as a stomach carcinogen, whereas only a few SD rats would be needed to give a positive result. Of course, the best result would be to use just SD rats, but there is no *a priori* way of deciding which strain is most likely to be sensitive. In the example given below, it is the SD strain of rats that are resistant, and it is clear that in the absence of any way of choosing a susceptible strain, it is best to split up the available material so as to increase the chances of having a few highly sensitive animals.

As a further example, consider Shellabarger et al.'s (6) study of the carcinogenicity of DES. They found that Sprague-Dawley rats failed to develop any excess tumors when treated with DES. However,

Table 2. Sample sizes needed to compare a treated and control group

True incidence in controls	True incidence in treated group	Sample size[a]
0.01	0.1	108
0.01	0.2	43
0.01	0.3	25
0.01	0.5	12
0.1	0.2	217
0.1	0.3	67
0.1	0.5	21
0.1	0.7	10

[a]In each group, assuming a one-tailed test with $\alpha = 0.05$ and a power $(1-\beta)$ of 0.9.

about 72% of treated ACI rats developed tumors compared with less than 1% in controls (the data have been averaged across an irradiation treatment).

Assume that on average half of all rat strains are like ACI and half are like SD, and one strain with 48 animals per group is chosen at random for the test. In this case there will be a 50% chance that a resistant strain of rats is chosen, and DES will be declared a noncarcinogen. Suppose alternatively that the test is set up using 12 rats from each of four strains chosen at random. The chance that at least one strain is susceptible will be $1 - (1/2)^4 = 93.75\%$. If only a single strain is susceptible, and 72% of rats of the susceptible strain develop tumors (between 8 and 9 tumors in the 48 treated animals), this would be judged to be statistically highly significant, and DES would be declared to be a carcinogen. The high strain sensitivity more than compensates for the reduced sample size of the individual strain (though total group sizes remain the same). So, by increasing the number of strains from one to four, the chance of detecting DES as a carcinogen would be increased from about 50% to over 90%. If eight strains were to be used, the chance that at least one was susceptible would be increased to 99.6%, but in this case if only 72% of the rats were to develop tumors, only just over four tumors would be expected, which would not be statistically significantly different from 0 in the controls. However, the chance that at least two strains were susceptible would be 96.5%, and with two susceptible strains the expected number of tumors would be over eight, giving a high chance of a significant difference. Thus, the use of eight strains would be marginally more powerful than the use of fours strains, though it may be impractical to use so many strains.

There is, however, a problem with the perception of the concept of sample size with these proposals. Take the case of the 4-strain experiment with 12 animals per strain. Many people assume that these proposals suggest a reduction in sample size from 48 animals to 12 animals per group. This is not the case. The treated and control groups would still each consist of 48 animals. It is only the arrangement of genotypes within that fixed sample size that is being discussed here. Theoretically, it would be possible to design the experiment with 48 strains with one animal of each strain in the treated and control group. Clearly, if this were perceived to be an experiment with a group size of one, it would be written off as being impossible to analyze statistically. In fact, a protocol of this sort is entirely analogous to the use of 48 sets of identical twins, and nobody would doubt that this could be analyzed statistically.

As noted above, and Felton and Gaylor (*28*) concluded that, on average, designs of this sort are likely to be statistically more powerful than the present design. Haseman and Hoel (*46*) reached somewhat similar conclusions when comparing the use of several isogenic strains with a pseudo-outbred stock assumed to be a random mixture of individuals from the isogenic strains. Their results suggest that the use of several isogenic strains would be better than the use of a single outbred stock. Lovell (*26*) found that such designs would be expected to detect more carcinogens and should provide more information on toxicological mechanisms which would be of qualitative value in assessing human risk, though the multistrain design would not improve methodology for extrapolating effects to low dose levels, except to the extent that it would avoid more false negative results. It has been suggested that the same benefits could be obtained by using another species. However, it would probably be impossible to add a third species without increasing the total number of animals used, and therefore the overall cost would rise substantially.

A Challenge for the NTP-CB

Part of the work of the NTB-CB is to conduct research into improved methods of screening for carcinogens. The case for using more than one strain has been carefully argued over a period of several years (*25,34–36*), and the statistical implications have been independently examined and found to be highly satisfactory. There appears to be scope for using strain differences in response as an additional tool for studying toxic mechanisms, though this requires more research. The continued use of a single isogenic strain with its associated failure to give any estimate of whether or not the same response would be observed with a different strain is open to criticism and may be leading to the establishment of a database on the carcinogenicity of chemicals which, far from being the "gold standard" against which short-term tests are evaluated, is actually hindering the development of new short-term tests. Green noted that

Failure to repeat short-term assays is inexcusable; failure to repeat rodent assays is entirely excusable, but it renders the data unsuitable as a reference baseline for evaluating other assays. An assay should not be used as a baseline, to test the predictivity of other assays, unless its ability to predict itself is known. (*47*: p. 369)

The design of the NTP-CB assay has remained almost unchanged for more than 20 years, though the science of toxicology has altered radically during this period, with increasing emphasis on short-term

tests and on understanding toxic mechanisms. This seems to be an appropriate time for scientists associated with the NTP-CB to initiate research into improving the design, particularly to increase the repeatability of the results. One avenue that clearly needs to be explored is the use of a multistrain design. It is impractical to repeat individual assays, but at least with a multistrain design there should be some estimate of the internal repeatabilty of the results. If the compound causes cancer in all or in none of the strains, it may be a more suitable chemical for validating short-term tests than if it causes cancer in only one organ of one strain.

Exactly how the other advantages and disadvantages of a multistrain design are to be evaluated needs detailed consideration. Failure to find strain differences in response to a few model carcinogens does not necessarily invalidate the design, as most potent carcinogens are likely to cause cancer in all species and strains, and in any case, the finding that a compound causes the same result in several strains is valuable information, as noted above. For general screening, the design is likely to be most helpful in detecting apparently weak carcinogens where in some strains increased latency to develop tumors and/or high sensitivity to acute toxicity means that most animals have not had time to develop tumors by the end of the test period. Table 1 gives an example where iron and hexachlorobenzene gave a high incidence of hepatocellular carcinoma in C57BL/10 mice (*48*) but none in DBA/2, possibly because mice of the latter strain died due to acute toxic effects.

One approach would be to use a multistrain assay for substances that have given equivocal results in a full-scale NTP-CB rodent assay and which need further investigation in any case. Responses to nongenotoxic carcinogens would be of particular interest. Maronpot et al. (*49*) found a number of substances that appeared to be carcinogenic in the strain A lung tumor assay but were negative in the NTP-CB, and there are numerous compounds that are negative in one species and positive in another (*50*). Some of these could be examined on a research scale rather than as a full rodent bioassay, possibly using quite small numbers of animals of several strains to see whether one or two strains could be found that are highly susceptible. It might be worthwhile doing lifetime studies on strains that develop high levels of preneoplastic foci in short-term studies, or small scale screening trials might be used to detect potentially sensitive strains before doing a full-scale bioassay. All sorts of strategies could be proposed, once the principle of using more than one strain was

accepted. If susceptible and resistant strains were found, then biochemical, pharmacological, and pathological differences between the strains could be examined to see if they are of any value in understanding mechanisms and improving available data for estimating human risk.

The genetic basis could also be examined in detail using molecular mapping methods which have only recently become available (*51*). Resistant and susceptible strains can be crossed to produce an F_2 hybrid or backcross segregating population, individuals of which are treated with the chemical. Each animal is then typed for simple-sequence length polymorphic (SSLP) loci covering all chromosomes. Any co-segregation between susceptibility and a particular SSLP marker implies genetic linkage. In this way, the genes contributing to strain differences can be mapped and, eventually, identified. Although this is a laborious procedure which could only be done for a few selected compounds, in the long run it could contribute enormously to an understanding of toxic and carcinogenic mechanisms. Devereaux et al. (*52*) and Festing et al. (*53*) have used these techniques to map the genes controlling sensitivity to the induction of lung tumors in mice, where strain A/J is highly sensitive and C57BL/6 is resistant. Susceptibility is controlled by the sex of the mouse and by loci on chromosomes 6, 9, 17, and 19, though the locus on chromosome 6 closely linked to or identical to the K-*ras*-2 genetic locus accounted for 60% of the total variation in one of the studies (*53*).

The challenge facing the NTP-CB now is to demonstrate the extent to which the assay results obtained with F344 rats and B6C3F$_1$ mice can be generalized to mice and rats with other genotypes and can be used to validate short-term assays. A first step would be to investigate the extent to which the NTP-CB results are strain-dependent.

REFERENCES

1. Ashby J, Morrod RS. Detection of human carcinogens. Nature 353:185–186 (1991).
2. Weideman M. Toxicity tests in animals: historical perspectives and new opportunities. Environ Health Perspect 101:222–225 (1993).
3. Ashby J, Purchase IFH. Will all chemicals be carcinogenic to rodents when adequately evaluated? Mutagenesis 8:489–493 (1993).
4. Omenn GS. Future research directions in cancer ecogenetics. Mutat Res 247:283–291 (1991).
5. Kihara M, Kihara M, Noda K. Lung cancer risk of GSTM1 null genotype is dependent on the extent of tobacco smoke exposure. Carcinogenesis 15:415–418 (1994).
6. Shellabarger CJ, Stone JP, Holtzman S. Rat differences in mammary tumor induction with estrogen and neutron irradiation. J Natl Cancer Inst 61:1505–1508 (1978).

7. Grasso P, Hardy J. Strain difference in natural incidence and response to carcinogens. In: Mouse hepatic neoplasia (Butler WH, Newberne PN, eds). New York:Elsevier, 1975;111–132.

8. Festing MFW. Inbred strains in biomedical research. London:Macmillan Press, 1979.

9. Kalter H. Dose–response studies with genetically homogeneous mice as a teratology testing and reassessment procedure. Teratology 24:79–86 (1981).

10. Holson JF, Gaines TB, Nelson CJ, LaBorde JB, Gaylor DW, Sheehan DM, Young JF. Developmental toxicity of 2,4,5-trichlorophenoxyacetic acid (2,4,5-T). I. Multireplicated dose–response studies in four inbred strains and one outbred stock of mice. Fundam Appl Toxicol 19:286–297 (1992).

11. Pohjanvirta R, Juvonen R, Karenlampi S, Raunio H, Tuomisto J. Hepatic Ah-receptor levels and the effect of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) on hepatic microsomal monooxygenase activities in a TCDD-susceptible and -resistant rat strain. Toxicol Appl Pharmacol 92:131–140 (1988).

12. Lijinsky W, Thomas BJ, Kovatch RM. Differences in skin carcinogenesis by methylnitrosourea between mice of several strains. Cancer Lett 61:1–5 (1991).

13. Nebert DW, Gielen JE. Genetic regulation of aryl-hydrocarbon hydroxylase induction in the mouse. Fed Proc 31:1315–1325 (1972).

14. Drinkwater NR, Ginsler JJ. Genetic control of hepatocarcinogenesis in C57BL/6J and C3H/HeJ inbred mice. Carcinogenesis 10:1701–1707 (1986).

15. Hanigan MH, Kemp CJ, Ginsler JJ, Drinkwater NR. Rapid growth of preneoplastic lesions in hepatocarcinogen-sensitive C3H/HeJ male mice relative to C57BL/6J male mice. Carcinogenesis 9:885–890 (1988).

16. Weindruch R, Masoro EJ. Concerns about rodent models for aging research. J Gerontol 46:B87–88 (1991).

17. Sontag JM, Page NP, Saffiotti U. Guidelines for carcinogen bioassay in small rodents. National Cancer Institute carcinogenesis technical report series no 1. DHEW Publication no (NIH) 76-801. Bethesda, MD: National Cancer Institute, 1976.

18. Festing MFW. Genetic factors in toxicology: implications for toxicological screening. CRC Crit Rev Toxicol 18:1–26 (1987).

19. Arcos JC, Argus MF, Wolf G. Chemical induction of cancer. London:Academic Press, 1968.

20. Festing MFW. Genetic variation in outbred rats estimated from DNA fingerprints: implications for toxicological screening (abstract). Hum Exp Toxicol 11:590–591 (1992).

21. Kuhnlein U, Zadworny D, Dawe Y, Fairfull RW, Gavora JS. Assessment of inbreeding by DNA fingerprinting: development of a calibration curve using defined strains of chickens. Genetics 125:161–165 (1990).

22. Rice MC, O'Brien SJ. Genetic variance of laboratory outbred Swiss mice. Nature 283:157–167 (1980).

23. Lyon MF, Searle AG, eds. Genetic variants and strains of the laboratory mouse, 2nd ed. Oxford:Oxford University Press, 1989.

24. Hedrich HJ, ed. Genetic monitoring of inbred strains of rats. Stuttgart:Gustav Fischer Verlag, 1990.

25. Festing MFW. Genetic variation in outbred rats and mice and its implications for toxicological screening. J Exp Anim Sci 35:210–220 (1993).

26. Lovell DP. Impact of pharmacogenetics on toxicological studies. Statistical implications. J Exp Anim Sci 35:259–281 (1993).

27. Falconer DS. Introduction to quantitative genetics, 2nd ed. London:Longman, 1981.

28. Felton RP, Gaylor DW. Multistrain experiments for screening toxic substances. J Toxicol Environ Health 26:399–411 (1989).

29. Tennant R. Stratification of rodent carcinogenicity bioassay results to reflect relative human hazard. Mutat Res 284:111–118 (1993).

30. Heston WE. Genetic aspects of experimental animals in cancer research. In: Experimental animals in cancer research (Muhlbock O, Nomura T, eds), Gann monograph 5. Tokyo: Maruzen, 1968;3–15.

31. Ridder GM, von Barge EC, Alden CL, Parker RD. Increased hyalin droplet formation in male rats exposed to decalin is dependent on the presence of α_{2u}-globulin. Fundam Appl Toxicol 15:732–743 (1990).

32. Carthew P, Rich K, Martin EA, DeMatteis F, Lim C-K, Yuan S-X, Manson MM, Festing MFW, Gant T, Smith LL, White INH. Differences in response of three rat strains to dietary tamoxifen: correlation of liver tumour formation with DNA damage, metabolism and cell proliferation (in preparation).

33. Smith LL, White I, Gant T. Increased clearance of tamoxifen and its metabolites in rat liver correlates with an increased expression of P-glycoportein. Presented at the Society of Toxicology Annual Meeting, Dallas, Texas, 1994.

34. Festing MFW. A case for using inbred strains of laboratory animals in evaluating the safety of drugs. Food Cosmet Toxicol 13:369–375 (1975).

35. Festing MFW. Properties of inbred strains and outbred stocks with special reference to toxicity testing. J Toxicol Environ Health 5:53–68 (1979).

36. Festing MFW. Inbred strains and the factorial experimental design in toxicological screening. In: Proceedings of the seventh international Council for Laboratory Animal Science symposium, Utrecht, the Netherlands (Spiegel A, Erichsen S, Sollveld HA, eds). Stuttgart: Gustav Fischer, 1980; 59–66.

37. Heston WE, Vlahakis, G Desmukes, B. Effects of the antifertility drug Enovid in five strains of mice, with particular regard to carcinogenesis. J Natl Cancer Inst 51:209 (1973).

38. Wolff GL, Gaylor DW, Blackwell B-N, Moore GE. Bladder and liver tumorigenesis induced by 2-acetylaminofluorine in different F1 mouse hybrids: variation within genotypes and effects of using more than one genotype on risk assessment. J Toxicol Environ Health 33:327–348 (1991).

39. Munro A, Davies TS. High dose levels are not necessary in rodent studies to detect human carcinogens. Cancer Lett 75:183–194 (1993).

40. Rao GN, Birnbaum LS, Collins JJ, Tennant RW, Skow LC. Mouse strains for chemical carcinogenesis studies: overview of workshop. Fundam Appl Toxicol 10:385–394 (1988).

41. Sprott RL. Mouse and rat genotype choices. Aging Clin Exp Res 5:249–252 (1993).

42. Sprent P. Applied nonparametric statistical methods. London:Chapman and Hall, 1993.

43. Piantadosi S. Clinical trials design program. Ferguson, MO:Biosoft, 1990.

44. Snedecor GW, Cochran WG. Statistical methods, 7th ed. Ames:Iowa State University Press, 1980.

45. Tatematsu M, Aoki T, Inoue T, Mutai M, Furihata C, Ito N. Coefficient induction of pepsinogen 1-decreased pyloric glands and gastric cancers in five different strains of rats treated with N-methyl-N'-nitro-nitrosoguanidine. Carcinogenesis 9:495–498 (1988).

46. Haseman JK, Hoel DG. Statistical design of toxicity assays: role of genetic structure of test animal population. J Toxicol Environ Health 5:89–101 (1979).

47. Green MHL. Short-term tests and the myth of the non-clastogenic mutagen. Mutagenesis 3:369–371 (1988).

48. Smith AG, Cabral JRP, Carthew P, Francis JE, Manson M. Carcinogenicity of iron in conjunction with a chlorinated environmental chemical, hexachlorobenzene, in C57BL/10ScSn mice. Int J Cancer 43: 492–496 (1989).

49. Maronpot RR, Witschi HP, Smith LH, McCoy JL. Recent experience with the strain A mouse pulmonary adenoma bioassay. Environ Sci Res 27:341–349 (1983).

50. Byrd DM III, Crouch EAC, Wilson R. Do mouse liver tumors predict rat tumors? A study of concordance between tumors induced at different sites in rats and mice. Prog Clin Biol Res 331:19–41 (1990).

51. Dietrich W, Katz H, Lincoln SE, Shin H-S, Friedman J, Dracopoli NC, Lander E. A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423–447 (1992).

52. Devereux TR, Wiseman RW, Kaplan N, Foley JF, White CM, Anna C, Watson MA, Patel A, Jarchow S, Maronpot RR, Anderson MW. Assignment of a locus for murine pulmonary adenoma susceptibility to proximal chromosome 19. Mammal Genome (in press).

53. Festing MFW, Yang A, Malkinson AM. At least four genes and sex are associated with susceptibility to urethane-induced pulmonary adenomas in mice. Genet Res (in press).

54. Isaccs JT. Inheritance of a genetic factor from the Copenhagen rat and the suppression of chemically induced mammary adenocarcinogenesis. Cancer Res 48:2204–2213 (1988).

55. Shirai T, Nakamura A, Fukushima S, Yamamoto A, Tada M, Ito N. Different carcinogenic responses in a variety of organs, including the prostate, of five different rat strains given 3,2'-dimethyl-4-aminobiphenyl. Carcinogenesis 11:793–797 (1990).

56. Watanabe H, Nakagawa Y, Takahashi T, Ito A. Effect of age and strain on N-methyl-N'-nitro-N-nitrosoguanidine tumorigenesis in ACI and Wistar Furthe rats. Carcinogenesis 9:1317–1318 (1988).

57. Melhem MF, Rao KN, Kunz HW, Kazanecki M, Gill TJ III. Genetic control of susceptibility to diethylnitrosamine carcinogenesis in inbred ACP (grc+) and R16 (grc-) rats. Cancer Res 49:6813–6821 (1989).