Toxicology and Applied Pharmacology 243 (2010) 300–314

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning

Scott S. Auerbach [a], Ruchir R. Shah [b], Deepak Mav [b], Cynthia S. Smith [a], Nigel J. Walker [a], Molly K. Vallant [a], Gary A. Boorman [a], Richard D. Irwin [a,*]

[a] National Toxicology Program, National Institute of Environmental Health Sciences, NIH, RTP, NC 27709, USA
[b] SRA International, RTP, NC 27709, USA

ARTICLE INFO

Article history:
Received 4 September 2009
Revised 18 November 2009
Accepted 20 November 2009
Available online 11 December 2009

Keywords:
Toxicogenomics
Cancer
Liver
Prediction
Alkenylbenzene
Rat

ABSTRACT

Identification of carcinogenic activity is the primary goal of the 2-year bioassay. The expense of these studies limits the number of chemicals that can be studied and therefore chemicals need to be prioritized based on a variety of parameters. We have developed an ensemble of support vector machine classification models based on male F344 rat liver gene expression following 2, 14 or 90 days of exposure to a collection of hepatocarcinogens (aflatoxin B1, 1-amino-2,4-dibromoanthraquinone, *N*-nitrosodimethylamine, methyleugenol) and non-hepatocarcinogens (acetaminophen, ascorbic acid, tryptophan). Seven models were generated based on individual exposure durations (2, 14 or 90 days) or a combination of exposures (2 + 14, 2 + 90, 14 + 90 and 2 + 14 + 90 days). All sets of data, with the exception of one yielded models with 0% cross-validation error. Independent validation of the models was performed using expression data from the liver of rats exposed at 2 dose levels to a collection of alkenylbenzene flavoring agents. Depending on the model used and the exposure duration of the test data, independent validation error rates ranged from 47% to 10%. The variable with the most notable effect on independent validation accuracy was exposure duration of the alkenylbenzene test data. All models generally exhibited improved performance as the exposure duration of the alkenylbenzene data increased. The models differentiated between hepatocarcinogenic (estragole and safrole) and non-hepatocarcinogenic (anethole, eugenol and isoeugenol) alkenylbenzenes previously studied in a carcinogenicity bioassay. In the case of safrole the models correctly differentiated between carcinogenic and non-carcinogenic dose levels. The models predict that two alkenylbenzenes not previously assessed in a carcinogenicity bioassay, myristicin and isosafrole, would be weakly hepatocarcinogenic if studied at a dose level of 2 mmol/kg bw/day for 2 years in male F344 rats; therefore suggesting that these chemicals should be a higher priority relative to other untested alkenylbenzenes for evaluation in the carcinogenicity bioassay. The results of the study indicate that gene expression-based predictive models are an effective tool for identifying hepatocarcinogens. Furthermore, we find that exposure duration is a critical variable in the success or failure of such an approach, particularly when evaluating chemicals with unknown carcinogenic potency.

Published by Elsevier Inc.

Introduction

The purpose of toxicity/carcinogenicity testing in rodents is to identify agents that may pose a carcinogenic risk to humans (Bucher and Portier, 2004). The current protocol used by the National Toxicology Program involves exposing a total of 800 rodents to 4 different levels of a test article for a duration of 24 months. Following the 2-year exposure period, a comprehensive histopathological assessment is performed on over 40 organs/tissues in all animals used in the study. All lesions identified by the primary pathologist are extensively reviewed by a panel of veterinary pathologists. Any calls related to carcinogenic activity are then evaluated by an independent panel of both private and public sector scientists with expertise in the area of toxicology (Chhabra et al., 1990). The high sensitivity of these studies makes them the current standard for identifying chemicals that pose a carcinogenic risk for humans (Huff, 1998).

The high sensitivity of the NTP carcinogenicity bioassay comes with significant cost in terms of money, time, animals and chemical. A 2-year bioassay can cost several millions of dollars and take up to 5 years to complete. In approximately 30 years since the inception of the bioassay only 1485 chemicals have been assessed (Gold et al., 2005). Currently there are over 75,000 chemicals on the US EPA's Toxic Substances Control Act Inventory (USEPA, 2004), an estimated 30,000 chemicals in widespread commercial use in the United States and Canada (Muir and Howard, 2006) and over 140,000 substances registered by the REACH (REACH, 2008). Only a small fraction of these agents have undergone carcinogenicity testing (Judson et al., 2009).

Abbreviations: RFE, (recursive feature elimination); support, vector machine (SVM).

* Corresponding author. Fax: +1 919 541 4255.
E-mail address: irwin@niehs.nih.gov (R.D. Irwin).

0041-008X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.taap.2009.11.021

Characterizing the carcinogenic activity of each of the untested chemicals using the 2-year bioassay is not a viable approach especially considering that some of the chemicals will need to be assessed individually and as mixtures. In light of these issues clearly more efficient methods need to be developed to identify chemicals that pose a carcinogenic risk.

Due to the combination of a broad landscape of untested chemicals in commerce and the current limitations of the bioassay there has been no shortage of attempts to identify methods that will allow for more rapid identification of potential human carcinogens. Efforts have ranged from purely computational SAR analysis (Benigni et al., 2007) to a range of biological approaches including various bacterial (Tennant et al., 1987) and mammalian cell-based in vitro genotoxicity assays (Isfort et al., 1996; Kerckaert et al., 1996), in vivo genotoxicity (Sasaki et al., 2000; Parry et al., 2002), mechanistic assessments based on receptor activation (Van den Berg et al., 1998), assessment of preneoplastic lesions (Elcombe et al., 2002; Ito et al., 2003; Allen et al., 2004), use of genetically modified animals (Eastin et al., 2001; Storer et al., 2001; Usui et al., 2001; van Kreijl et al., 2001; Lambert et al., 2005), and approaches that combine a number of these technologies (Benigni and Zito, 2004; Cohen, 2004). Most have either fallen short in their predictive ability or have not undergone extensive validation due to the expense of such a project (Jacobs, 2005). Some of the reasons for failure of the predictive strategies range from inadequate training data in the case of in silico predictive models to the inadequacy of the test systems to address certain modes of action (e.g. transgenic mice) (Bucher and Portier, 2004).

Genomic technology allows a researcher to query, in exquisite detail, molecular level changes in biology. When used in combination with machine learning, genomics has excelled in determining cancer diagnosis, prognosis and chemotherapeutic response (Garman et al., 2007). Recently, a number of groups have taken advantage of this technology to build carcinogenicity prediction models from gene expression data (Kramer et al., 2004; Nie et al., 2006; Fielden et al., 2007; Thomas et al., 2007; Ellinger-Ziegelbauer et al., 2008; Uehara et al., 2008). Most of these studies have focused on liver because it is common target for chemical-induced carcinogenic transformation and predictive models have the potential to accelerate carcinogenicity hazard characterization. For comparison purposes the details of these studies are reviewed in the SupplementaryIntroduction. Overall these studies demonstrate the success and utility of such an approach. In addition, the studies that have incorporated a consideration of hepatocarcinogenic mode of action indicate that it is possible to differentiate between genotoxic and non-genotoxic modes of action which is useful for determining human cancer risk and is not currently possible with the traditional in vivo toxicity/carcinogenicity assessment methods.

The toxic effects of a chemical are both a function of dose and duration of exposure (Rozman, 2000). Most of the genomic studies described above focused on identifying liver carcinogenicity signatures from animals exposed to chemical for 28 days or less, a duration that is referred to in traditional toxicology as subacute. We hypothesized that exposures up to 90 days would accentuate the expression of genes related to carcinogenic activity and therefore allow the models to achieve a higher degree of certainty when making predictions. Furthermore, we reasoned that longer exposure durations would limit the influence of mode of action genes and allow for better identification of predictive genes with biology related to processes involved in the formation of neoplasms that are typically manifest secondary to the primary toxicity. We feel that the idea of a shared precancerous biology (that is independent of a specific chemical challenge) is not unreasonable since the process of cancer manifestation is a continuum and most cancers share a degree of universal biology that is manifest in their gene expression (Whitfield et al., 2006). To test this hypothesis we exposed male F344 rats to a collection of structurally diverse hepatocarcinogens and non-hepatocarcinogens for 2, 14 or 90 days (carcinogen vs. non-carcinogen study (CVNC), performed genome-wide hepatic gene expression using Agilent 4X44K microarrays (41,000+ rat genes and transcripts) and created models that identified chemicals with hepatocarcinogenic activity. We then independently validated these models using hepatic expression data derived from rats exposed to a collection of alkenylbenzenes (flavoring agent study (FA)). Alkenylbenzenes are food additives with a range of hepatocarcinogenic properties (Miller et al., 1983). In the course of these studies we specifically address the effect of dose level and exposure duration on classification accuracy, in addition to evaluating the molecular biology that is associated with carcinogen exposure.

Materials and methods

Chemicals used for dosing. All chemicals administered to rats in this study are listed in Table 1. Pre-start chemistry assessments indicated that all chemicals were at least 98% pure. All feed formulations underwent homogeneity assessments. All dose formulations were within 10% of the target concentration throughout the study.

Chemical diversity analysis. Leadscope Enterprise 2.4.15-6 (Leadscope Inc., Columbus, OH) was used to evaluate the structural similarity between the 13 chemicals used in the study. The Tanimoto distance was calculated using the chemical figure print derived from the 2-dimensional mol file of each chemical. Leadscope uses a 27,000+ feature set to derive a chemical figure print. This feature set is much larger than is used by most applications and causes the Tanimoto distances to be smaller relative to those calculated by other structural analysis programs.

Animal treatments and tissue collection. All animal studies were performed at Battelle (Columbus, OH) under the direction of Milton Hejtmancik, Ph.D., D.A.B.T. and Laurene Fomby, D.V.M., Ph.D., D.A.B.T. Male F344/N rats approximately 8 to 10 weeks old were obtained from Taconic Farms (Germantown, NY). After a 10 to 14-day quarantine/acclimation period, the rats were randomly assigned to treatment and control groups. The light/dark cycle was 12 h on/12 h off with the lights coming on at 6 AM and going off at 6 PM. Rats were housed 3 per cage with *ad libitum* access to NTP 2000 feed and city water. In the CVNC study groups of 24 male rats were administered chemical either in feed, in drinking water, or by gavage (Table 1). On days 3 (2 nights of exposure), 15 (14 nights of exposure) and 91, (90 nights of exposure), 6 rats from each chemical group plus 6 rats from the appropriate control group were necropsied between 8 and 10 AM. Detailed information on the number of animals in each treatment group can be found in Table 1. In the alkenylbenzene flavoring agent (FA) study each chemical was given at 2 dose levels: 0.2 mmol/kg/day (low dose (L)) and 2.0 mmol/kg/day (high dose (H)) by corn oil gavage 5 days per week. Each dose group and the appropriate control consisted of six animals in the 2 and 14-day studies and 10 animals in the 90-day study. Necropsies were performed 24 h after the last administered dose. At necropsy rats were anesthetized with isofluorene, blood drawn for clinical chemistry via cardiac puncture, the left and median lobes of the liver removed, and animals euthanized by exsanguination. A cross-section of each lobe was obtained for histopathology. The remainder of the left and median lobes of the liver were minced quickly into very small pieces and dropped in liquid nitrogen within 4 min of euthanasia and stored at −80 °C.

Hematology and clinical chemistry. Blood collected from all animals in both studies was analyzed for routine hematology and clinical chemistry markers including erythrocyte count, mean corpuscular volume, hemoglobin, packed cell volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, erythrocyte morphologic assessment, leukocyte count, leukocyte differential,