Abstract. The identification of cancer biomarkers that will predict susceptibility to disease and subsequent clinical outcome are key components of future genomics-based tailored medical care. Animal models of disease provide a rich resource for the identification of potential cancer biomarkers. Animal models of prostate cancer in particular offer the potential to identify cancer genes associated with dietary and environmental factors. The key issue is the timely and efficient identification of candidate genes that are likely to impact on human prostate cancer. Here, we demonstrate comparative genomics-based methods for the identification of candidate genes in animal models that are associated with human chromosomal regions implicated in prostate cancer. Using publicly available bioinformatics tools, comparisons can be made between cancer-specific datasets, genomic sequencing data and cross-species comparative maps to identify potential cancer biomarkers. This process is demonstrated by using rat models of prostate cancer to identify candidate human prostate cancer genes. Genes identified through these techniques can be screened as biomarkers for response to chemopreventive agents, as well as being used in transgenic or knockout mice to engineer better animal models of human prostate cancer. The bioinformatics techniques outlined here can be used to leverage genomic data from any animal cancer model for use in the study and treatment of human cancer.

The Use of Comparative Mapping in Gene Identification

The genomic sequencing of multiple species has led to extensive studies in comparative genomics (1-3). These studies have identified the relationship between genomes and defined syntenic regions for genomic similarity (4), thus laying the groundwork for their use in cancer research. Initial cancer studies have utilized comparative genomics to identify orthologs of cancer genes (5, 6), as exemplified for the Wnt signaling pathway (7, 8). Additional efforts have focused on the identification of syntenic regions for genetic mapping data in neuroblastoma and lung cancer animal models (9-11). Recently, experimental comparisons have been made across species at the level of whole transcriptomes using gene orthologs. Particular examples of this work include the analysis of mouse and human hepatocellular carcinoma (12, 13) and our own work in rat and human prostate cancer (14). Thus, the ability to use bioinformatics processes and tools has made it possible for researchers to examine and evaluate their cancer-specific genomic data from animal models for relevance in human cancer. Through simple publicly available bioinformatics tools for visualizing genomic information (genome browsers), an investigator can examine their cancer-specific genetic regions, expression results, and chromosomal region data against the backbone of the rat, mouse, or human genomes and, through cross-species comparative genomics, identify cancer genes based on multiple lines of scientific evidence. These local datasets, once studied and...
published, can then be made publicly available, to be re-analyzed and expanded upon by subsequent scientists. Here, we outline a bioinformatics-based process in which datasets from animal models of prostate cancer can be examined and compared to human prostate cancer data using a Genome Browser. These datasets include rat prostate cancer microarray gene expression and genetic mapping studies, mouse transgenic and knockout data, rat mouse and human genomic sequence data, and comparative mapping data across the rat, mouse and human genomes. Through this process, comparative mapping can be used to identify syntenic regions of the rat genome that, in the human, are associated with prostate cancer risk. Within these regions, genes expressed in rat prostate tumors under specific environmental effects can be prioritized for validation studies.

Animal Models as a Source of Novel Cancer Genes

Animal models of cancer are important to the understanding of the pathophysiology of neoplasia and the development and testing of treatments. In particular, animal models are ideally suited for the examination of factors including diet, toxin exposures and drug responses, due to the ability to control the environmental exposures. This is important in the study of complex genetic traits such as the susceptibility to prostate cancer, with its associated environmental and genetic risks. Animal models of prostate cancer can be grouped into two general categories; i) genetically-engineered models, in which known cancer genes have been studied, and ii) spontaneous tumor models, where the genes involved in tumor risk are not known. Since our goal is the identification of novel cancer genes, we focused on the use of animal models in which the genes are not known. These types of studies can also be used in genetically-engineered animal models to identify additional genes that modify the tumor phenotype or outcome. For spontaneous models of prostate cancer, most work has been done on dog and rat models. These have been limited in their application to human prostate disease because of their high cost, low tumor incidence and dissimilarities to prostate anatomy (15, 16). Yet, rat models have shown value in the study of hormonal-based carcinogenesis and because they mimic human responses to chemotherapeutics and dietary factors in prostatic neoplasia (17-19). In addition, the development of mouse prostate cancer models, using transgenic and knockout technologies, allows for the genetic manipulation and combination of prostate cancer genes to produce better models. This is exemplified by the development and interbreeding of genetically-engineered mouse prostate cancer models such as the NKKX3.1, PTEN, p27 and Androgen receptor transgenic or knockout animals (20-25). Through the identification of additional genes in other animal models and their breeding into these mice, we can leverage the strength of the mouse model system to create better animal models for cancer research (26).

Genetic and Genomic Data in Animal Models of Cancer

Animal models offer the opportunity to collect large amounts of both genetic and genomic data associated with specific tumor characteristics. This has been demonstrated through the collection of gene expression microarray and genetic mapping data, both of which provide the opportunity to identify novel cancer genes. Genetic mapping in animal cancer models has been most extensively demonstrated using rat models of breast cancer (27-31). Prostate cancer genetic mapping studies are in their infancy, although preliminary mapping data has been generated for quantitative trait loci (QTLs) associated with prostate tumor formation and metastasis (Datta and Suckow, unpublished). Genomic studies have been more commonly performed and include studies of rat tumors under various conditions, for example aging and dietary changes (32-35). These datasets provide a rich source of candidate genes for subsequent validation.

Collecting Human Genomic and Genetic Cancer Data for Comparison

Animal data can be analyzed for relevance in human cancer by filtering it against human cancer genomic and genetic data. The sequencing of both the rat and mouse genomes allows for comparative mapping across species (2, 3). By linking the genomic cancer data across genomes, one can leverage syntenic mapping to prioritize genes found in animal studies for validation based on their location within chromosomal regions involved in human cancer. In order to accomplish this, one must have cancer-specific human genomic and genetic data. Specific sources include familial genetic mapping studies, cytogenetics and loss of heterozygosity studies, and comparative genomic hybridization and gene expression microarray studies. While some of these data types have been collected in central repositories such as a CGAP (36), much still resides in individual publications and laboratory files. We have developed a data set of human chromosomal regions associated with prostate cancer, named ChromSorter PC (37) (Figure 1). This is a hand-annotated collection of data extracted from publications identified in automated searches of Medline using the key words "human", "prostate cancer" and "chromosome". The literature was examined, sorted and annotated for prostatic intraepithelial neoplasia, prostate cancer and prostatic metastasis. In each article, data was extracted regarding the experimental methods used (comparative genomic hybridization, loss of heterozygosity,
karyotyping, chromosomal transfer), demographics (age groups, ethnicity) and chromosomal regions, including chromosomes, chromosomal arms and cytogenetic bands and, if available, genetic markers. These datasets, along with additional datasets for comparative genomic hybridization or prostate cancer gene expression studies, can be used as benchmarks for comparison with animal model data.

**Integrating Local Data Using Large Scale Genomic Datasets**

A crucial question is how specific cancer genes will be identified and sorted from the large quantities of genomic data. In this sorting process, bioinformatics-based comparative genomics will play a crucial role by associating genes with chromosomal regions that are linked to human cancer, thus making them strong candidates for further study. Bioinformatics and its associated computer-based tools allow for the integration and comparison of various types of genomic data including sequence, expression and functional information to cross entire genomes. The ability to rapidly organize this genomic data has been the hallmark of such bioinformatics and genomics pipelines as the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), the University of California at Santa Cruz genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and the Ensembl database (http://www.ensembl.org/). While

Figure 1. ChromSorter PC data for human chromosome 1p36. Data is displayed on a generic genome browser (http://www.gmod.org). Each referenced chromosomal data element is displayed with respect to the human genome, and is linked to the original article in PubMed. The syntenic regions in the rat genome have been selected and are displayed at the bottom.
extremely valuable in the analysis of genomic sequence data with respect to gene structure, function and even expression, the data presented is often very general and not focused on specific cancers. The true value of these Internet tools for cancer research is through the ability of users to upload their own locally generated cancer data for comparison with the stored genomic data. This provides a framework for comparing cancer data to known genes and
across genomes. The ability to upload and compare data is a feature that can be found on the UCSC genome browser. Some cancer-specific databases are now being developed that will allow for the analysis of cancer-specific data (CMAP, http://cmap.nci.nih.gov/), but currently do not have the ability to compare across species. For groups with biocomputing capabilities, the implementation and use of their own genome browsers allows for the collection of specific cancer-related data of relevance to the laboratory. This data can then be analyzed and reviewed without the need to reload datasets on a public genome browser. One particular tool we have used is the Generic Genome Browser developed by Lincoln Stein at Cold Spring Harbor Laboratories (http://stein.cshl.org/) and made publicly available through the Generic Model Organism Database Project (http://www.gmod.org/, Figures 1-3).

Using Cross-species Comparisons for Prostate Cancer Gene Identification

Animal data must be evaluated in the context of human cancer relevance. Thus, the identification and prioritization of genes identified in rat prostate cancer studies needs to be carried out in the context of human prostate cancer data. We have previously shown that gene expression profiles between rat and prostate cancer cells lines show a dramatic level of similarity (14). In addition, we have shown that, by concentrating on genes that demonstrate consistently similar changes to nutrients like Selenium across species, one can identify significant pathways involved in human prostate cancer (14). Here, we demonstrate how these studies can be extended through cross-species comparative mapping with genetic and genomic data. While demonstrated for prostate cancer, the techniques are applicable to any cancer system in which animal model data is available or can be generated.

Predicted Synteny Between Animal Model and Human Prostate Cancers

We have collected genetic data from a spontaneous rat model of prostate tumors. The Lobund-Wistar (L-W) rat is predisposed to spontaneously develop metastasizing tumors of the anterior prostate-seminal vesicle complex (38). While the anatomic similarity to human prostate cancer has been debated, the clinical similarities, including hormonal modulation, onset with age, similar chemosensitivity and dietary profiles, have been well established (38-41). As such, this model offers an opportunity to examine environmental effects in hormonal carcinogenesis. We have previously documented that the tumor susceptibility in the Lobund-Wistar rat can be transferred to a tumor-resistant Copenhagen rat strain through hybrid breeding, suggesting a genetic mechanism for tumor sensitivity (or resistance) (42). In a genetic characterization of the Lobund-Wistar colony using over 350 polymorphic genetic markers, it was noted that the animals in the Lobund-Wistar colony were 92% inbred over their genomes, with 8% of the genetic loci containing polymorphic (multiple) alleles. Using these regions with polymorphic markers, genotyping was performed on archival primary and metastatic tumors from over 50 Lobund-Wistar rats. In this genome sharing approach, we asked whether, in the polymorphic regions, one of the variant alleles preferentially segregated with tumor or metastasis formation. The results identified five chromosomal regions, two for tumor and three for metastasis. The regions identified include portions of rat chromosomes 1 and 18 for tumor formation, and areas on rat chromosomes 1, 8 and 10 for the metastatic spread of tumor. Comparative mapping between syntenic regions of the rat and human genomes reveal that these chromosomal regions in the Lobund-Wistar rat are syntenic to chromosomal regions involved in human prostate cancer. Using the human prostate cancer chromosomal data in the ChromSorter PC database (37), the chromosomal regions of the rat genome have predicted synteny to human chromosomes 1p36 (Figure 2) (43-45), 8q24 (46-50), 5q22-25 (51) and 6q22-27(Figure 3) (52-54), all areas that have been associated with prostate cancer by either familial genetic mapping, loss of heterozygosity studies and/or comparative genomic hybridization. These studies are currently being followed up by a full scale genetic backcross to better characterize the chromosomal regions.

Combining Synteny and Gene Expression Data to Leverage Animal Models for Prostate Cancer Gene Identification

In addition to these genetic studies, synteny data was used to overlay microarray gene expression data from rat prostate cancer cell lines treated with the chemopreventive agent Selenium on the human genome. In this way, the gene expression data from animal models can be sorted to prioritize genes of significance in human prostate cancer. The gene expression datasets were visualized across the rat and human genomes using a generic Genome Browser anchored to either human or rat genome sequence data. Visualization of the genomic data across species allows the syntenic alignment of data, thus leveraging the strengths of each species. Overlap was visually identified based on a concordance of gene expression microarray data and regions of synteny to human prostate cancer chromosomal regions. Examples of two genes are presented and include the endothelial-specific receptor tyrosine kinase Tie-2/Tek and the protein S100A4. Tie-2/Tek is present on human chromosome 9p21, a region involved in chromosomal aneuploidy in human prostate cancer (55, 56). This gene is
a receptor for angiopoietins and modulates the activity of vascular endothelial growth factor (VEGF) (57). It has been found to be elevated along with VEGF in the serum of prostate cancer patients (58, 59). An additional gene identified with respect to Selenium treatment is S100A4. This gene maps to Xq27-28, a region associated with human prostate cancer based on familial genetic mapping studies (60-63). S100A4 has recently been associated with prostate cancer (64, 65), along with multiple other members of the S100 protein family (66, 67). These genes, thus, become prioritized candidates for biomarker validation and use in the engineering of better animal models of prostate cancer.

**Visualizing the Genomic Similarities of Animal Models to Human Disease, Now Possible with Available Bioinformatics Tools**

Bioinformatics tools and processes demonstrate a new opportunity for scientists to perform cross-species syntenic analysis of their animal model genomic and genetic data to identify genes of significance in human cancer. Simple comparisons can be performed using Internet-enabled tools such as the UCSC Genome Browser, while local tools, such as the generic genome browser, are open source and can be modified and expanded by individual groups for more advanced comparative mapping studies. Through the publication and release of the associated data, further analysis can be achieved, thus advancing the field. These processes and the associated bioinformatics tools are freely available and can be applied to any disease process or animal model evaluation.

Using these tools, prostate cancer-specific genes from animal studies were selected as potential cancer biomarker candidates, based on their association with chromosomal region data in human prostate cancer. Thus, the process provides a novel method for identifying and prioritizing new cancer genes based on animal models. In this way, candidate gene identification can be accelerated. This data also demonstrates the applicability of prostate cancer animal models, in particular with dietary or chemoprevention studies, to identify genes related to nutrients and prostate cancer. This current method also has the capability of analyzing other animal models of human disease at the genomic level, identifying both common and differing features that indicate the strengths and weaknesses of the model. Subsequent selection of specific genes or regions for breeding and genetic manipulation will allow for the development of better animal models of cancer.

**Acknowledgements**

This work was supported in part from NCI R21CA098032 to M.W.D., M.A.S. and M.P. and a grant from the Elsa Pardee Foundation, USA, to M.P.

**References**


44 Berry R et al: Linkage analyses at the chromosome 1 loci 1q24-25 (HPC1), 1q42.2-43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer. Am J Hum Genet 66(2): 539-546, 2000.


