DNA integrity during the cell life requires the thorough understanding of the maintenance of genome stability. It contains extensive data on the nomenclature, ontology, structure and function of proteins related to the DNA integrity mechanisms such as chromatin remodeling, histone modifications, DNA repair and damage response from eight organisms: Homo sapiens, Mus musculus, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Escherichia coli and Arabidopsis thaliana. DNAtraffic contains comprehensive information on the diseases related to the assembled human proteins. DNAtraffic is richly annotated in the systemic information on the nomenclature, chemistry and structure of DNA damage and their sources, including environmental agents or commonly used drugs targeting nucleic acids and/or proteins involved in the maintenance of genome stability. One of the DNAtraffic database aim is to create the first platform of the combinatorial complexity of DNA network analysis. Database includes illustrations of pathways, damage, proteins and drugs. Since DNAtraffic is designed to cover a broad spectrum of scientific disciplines, it has to be extensively linked to numerous external data sources. Our database represents the result of the manual annotation work aimed at making the DNAtraffic much more useful for a wide range of systems biology applications.

INTRODUCTION

A comprehensive understanding of the maintenance of DNA integrity during the cell life requires the thorough characterization of many simple data concerning all nuclear processes involving DNA, and including replication, repair, recombination (3R) and transcription. The major processes that regulate chromatin structure and counterbalance its repressive effects are: (i) chromatin remodeling, (ii) post-translational histone modifications and (iii) histone replacement. Chromatin is a dynamic structure that modulates the access of regulatory factors to the genetic material. The main role of DNA molecules is the long-term storage of information, genetic instruction used in the development and functioning of all known living organisms (with the exception of RNA viruses). Cells are continuously exposed to damaging agents whose action results in modification of nucleic acids. DNA damage from endogenous sources gives rise to 20 000 lesions/mammalian cell/day (1). Lesions are also caused by errors in DNA metabolic processes, including the formation of single and double-strand breaks from the collapse of replication forks and the introduction of modified nucleic acid bases during DNA replication. Counting all together, daily the 10^16-10^18 repair events occur in a healthy adult man (10^12 cells) (2). On the other hand, DNA damage is also caused by the environmental factors such as chemicals, UV light and ionizing radiation. Also, DNA structure and some proteins involved in DNA replication and repair are targets for the drugs used during chemotherapy (3). The available anticancer drugs have distinct mechanisms of action, which may vary in their effects on different types of normal and cancer cells. Their role is to slow and hopefully halt the growth and spread of a cancer.

Across the evolutionary spectrum, living organisms depend on high-fidelity DNA replication and recombination mechanisms have to respond to DNA damage and balance between the harmful and beneficial effects of manipulation into the genetic code. The knowledge of the processes in charge of DNA metabolism is critical to our understanding of how and why the genome is affected during the lifespan of the organism, and how the DNA repair systems efficiently work via several different pathways to protect the genome from potential mutagenic...
modification and allow accurate transmission of genetic information (4). Unrepaired lesions or strand brakes left in DNA might be the result of dysfunction in DNA repair, and lead to aging, carcinogenesis or neurodegeneration (5,6). Some pathological disorders are directly related to defects in DNA repair, telomere maintenance or DNA damage response machinery (7–9). At the same time, defects in DNA repair, telomere maintenance or DNA machineries performing these tasks need to gain access to the DNA that is packaged into chromatin or nucleoid. The main aim of the DNATrffic database is to cover and elucidate the interdisciplinary knowledge linking all aspects of the DNA integrity processes (e.g. chromatin dynamics, DNA replication, damage signaling and DNA repair), DNA damage and drugs interacting with DNA or proteins directly enrolled in DNA metabolism and connect all pieces together for the coordination of steps within a pathway or for crosstalk between different pathways. As transcription, recombination and DNA integrity are central components in the evolution of recent genome structures, and because replication, recombination and repair (3R) were fundamental prerequisites for the origin of life, all these topics are taken under analysis and serve as the cohesive force underlying this comprehensive DNA topic-focused database (18).

PathCARD
We used KEGG (13) and Reactome (12) databases for data implementation about pathways and networks concerning DNA metabolism. Some data like prokaryotic SOS response and translesion synthesis (TLS) were directly added by our DNATrffic team. All proteins are classified according to the orthology class, and next to the DNA integrity networks: chromatin organization and histone modifications, replication, damage checkpoint, DNA repair, modulation of nucleotide pools and so on (Table 1). It must be emphasized that all described processes are tightly connected to each other and they act in concert sharing some steps and/or proteins. Known functions of proteins are indicated in the curator comments section of each entry. A special emphasis is devoted to the function of that protein within DNA metabolism pathways but we also refer to alternative roles in other pathways. Additionally, all Gene Ontology terms associated to that protein are listed. The pathway in which a given protein is playing a role is also explored by linking from DNATrffic to the pathways included in the KEGG and Reactome databases.

ProteinCARD
According to the DNA metabolism network we used the UniProt (19), KEGG (13) and National Center for Biotechnology Information (NCBI) databases for protein data implementation into DNATrffic database for eight model organisms commonly used for DNA study. We collected 2921 proteins, for example—582 for Homo sapiens, 277 for Saccharomyces cerevisiae and 91 for E. coli (as of 13 October 2011). Using direct access from DNATrffic to protein all users can obtain unusual view of well-known proteins from model organisms but classified into the orthology classes. This innovation may be useful for the systems biology research and proper selection of the model organism for further study (Figure 1) of selected pathway. Amino acids and DNA sequences were downloaded from Ensemble. When available, links to the protein 3D structure in Protein Data Bank (PDB) were provided and 2D picture is visible in the single ProteinCARD entry. If annotated, possible physical interactions with other proteins were obtained through IntAct.

DETAILS RELATING TO DNATrffic’s OVERALL DESIGN AND DATA STRUCTURE DEPICTION CONVENTIONS
The aspects of the biochemistry and molecular biology of the genome dynamics during the cell life are the key for learning genome stability networks. During DNA replication, transcription and DNA repair, the cellular machineries performing these tasks need to gain access to the DNA that is packaged into chromatin or
implemented from OMIM (22) and KEGG databases (13) as well as directly from PubMed. This action needed manual annotation work. Each disease possesses its own DiseaseCARD entry with a succinct description, link to protein(s) and sometimes the picture of the symptoms. Reciprocal links to diseases are also available in each protein and pathway field (Figure 1).

**DamageCARD**

As of 13 October 2011, we collected information about 146 different types of damage in the DNA. Many of them describe general classes of damage events such as methylation or oxidative damage, or single-strand breaks or base loss, which are independent of the local sequence. About 50 chemical compounds that cause DNA damage were connected to the appropriate types of damage. Each type of damage is described on its own DamageCARD entry that includes information about the potential source (e.g. spontaneous formation, intermediate in some DNA repair process, methylating agents, etc.), proteins that may recognize its presence in the DNA, keywords that facilitate analyzing its context and external links (if available) to: PubChem Compound (CID), PubChem Substance (SID), ChemSpider, KEGG Compound, ChEBI and ChEMBL. DNAtraffic database also displays the unique chemical structures of DNA lesions in 2D and provides atomic coordinates for download in the smiles, InChi and InChiKey format.

**DrugCARD**

Till now, we collected information about over 181 different types of drugs interacting with DNA or proteins involved in nucleic acids metabolism. Data were implemented from DrugBank, T3DB, Therapeutic Target Database (TTD), KEGG Compounds databases (13,23–25). Each type of drug is described on its own DrugCARD entry that includes information about the potential application (e.g. anticancer treatment, DNA topoisomerase inhibitor and other), drug–protein or drug–DNA interaction and external links to DrugBank, KEGG Compound, PubChemCompound, PubChem Substance, ChemSpider, ChEBI, ChEMBL and TTD databases. DNAtraffic database also displays the unique chemical structures of drugs in 2D and provides atomic coordinates for download in the smiles, InChi and InChiKey format.

**SCHEME OF THE DNAtraffic DATABASE ARCHITECTURE**

The unordered data are difficult to interpret and many of the connections are lost. The OWT ontology provides the clear view and discovers the new connections. DNAtraffic database has been implemented using the Django web framework (http://www.djangoproject.com/). It uses a PostgreSQL relational database to store data (http://www.postgresql.org/). Scripts are written in Phyton language. DNAtraffic database is freely available and can be accessed at http://dnatraffic.ibb.waw.pl/dnatraffic/.
EXPANDED DATABASE LINKAGES

Because DNAtraffic was designed to cover a broad spectrum of scientific disciplines, it must be extensively linked to many external databases. Until now, DNAtraffic contains up to 15 database hyperlinks including links to KEGG (13), UniProt (19), OMIM (22), PDB (26), PubChem (27), ChEBI (28), ChEMBL, GenBank (29), Pfam (30), GeneCards (31), GenAtlas (32), HGNC, PubMed, ChemSpider (33) and TTD (25).

CONCLUSION

Researchers of the various chromatin structure and DNA repair processes have recently embraced approaches in which global measurements of gene expression and the proteome can be combined with genome-wide screening of sensitivity mutants to develop an integrated view of how cells respond to and protect themselves against DNA damaging agents. The emerging picture from these global genomic studies is quite different from the previous concept of DNA repair, cell cycle control and induction of...
apoptosis as being independent processes. In fact these processes appear to form a fully integrated network. Integration of these genome-wide measurements allows the development of specific models of response networks that could not have been detected or discerned previously.

DNAtraffic database is the first platform for systems biology of DNA integrity during the cell life, and can be also integrally involved in translational research (18). This includes the identification of small molecule inhibitors of novel DNA damage response (DDR) pathways that put new light on the causes of cancer or have potential uses in treatment.

DNAtraffic contains a significant number of data. As highlighted throughout this article, numerous improvements have been made in the quantity, quality, depth and organization of the information provided. DNAtraffic contains illustrated DNA networks in the cell, protein, damage and drug structures data and pictures. DNAtraffic also offers expanded database links. It is hoped that DNAtraffic will continue to develop to fulfil the needs of its users and provide an increasingly useful, information-rich DNA metabolism resource.

ACKNOWLEDGEMENTS
The authors would like to thank Professor Jaroslaw Kusmierak (Institute of Biochemistry and Biophysics, Polish Academy of Sciences) for his help with the DNA damage nomenclature and chemical structure drawings.

FUNDING
A grant from the Norwegian Financial Mechanism (PNRF-143-AI-1/07, in part); Polish Ministry of Science and Higher Education (N N301 165835, in part). Funding for open access charge: partially waived by Oxford University Press (55%), and Institute of Biochemistry and Biophysics PAS, Warsaw, Poland (45%).

Conflict of interest statement. None declared.

REFERENCES


