

Commentary

Pokemon expression in malignant glioma: an application of bioinformatics methods

RICHARD A. ROVIN, M.D., AND ROBERT WINN, PH.D.

Division of Neurosurgery, Marquette General Hospital; Upper Michigan Brain Tumor Center; and Department of Biology, Northern Michigan University, Marquette, Michigan

Object. In this report the authors review the role of bioinformatics in the design of a research project in which the molecular genetics of malignant glioma were studied. A project to characterize *Pokemon* expression in malignant glioma was developed, refined, and implemented using bioinformatics methods.

Methods. Using the resources available from the National Center for Biotechnology Information, the messenger RNA (mRNA) sequence for *Pokemon* was determined. With this information and online primer design tools, novel primers were designed that would specifically amplify *Pokemon* mRNA by using reverse transcription–polymerase chain reaction assays.

Conclusions. The promise of bioinformatics is in the rapid and widespread dissemination and analysis of genomic information. This information is then used in research investigating the genetic basis of disease. In this paper the authors review the bioinformatics methods used in their study of *Pokemon* expression in malignant glioma.

KEY WORDS • *Pokemon* gene • malignant glioma • bioinformatics • p14ARF inactivation

Bioinformatics has been variously defined, although in general it includes the mathematical modeling of biological processes and molecular structures as well as the creation of databases in which to store biological information, especially genetic data (see <http://bioinformatics.org/faq/>). Undoubtedly, the greatest achievement involving bioinformatics to date has been the Human Genome Project, which was completed in 2003. Now that we have entered the postgenomic era, bioinformatics methods are shifting from data acquisition to data analysis. This is in keeping with the National Human Genome Research Institute's vision of genomics research: "translating the information and resources generated by the Human Genome Project into medical value" (<http://www.genome.gov/11006873>).

Beginning in the fall of 2004, plans were made to develop a Brain Tumor Center in Upper Michigan. In addition to clinical treatment and patient advocacy programs, we wanted to develop a basic science research program to study the molecular genetics of malignant gliomas to guide treatment.

In January 2005, the protooncogene *Pokemon* was described.² *Pokemon* represses the transcription of the tumor suppressor p14ARF in humans (p19ARF in mouse), and

has been shown to be highly expressed in human lymphomas and in breast, lung, colon, prostate, and bladder carcinomas. *Pokemon* expression in malignant glioma has not been described.

We hypothesized that *Pokemon* expression was important in gliomagenesis. The following points, identified through a bioinformatics process, support our hypothesis and led to our study of *Pokemon* expression in adult human glioma.

- *ZBTB7A* is the gene coding for *Pokemon*.
- *ZBTB7A* is located at chromosome 19p13.3.
- Comparative genomic hybridization analysis of gliomas revealed high-level gains of the 19p13.3 locus in GBM tissue.⁶
 - The 19p gain is seen in 50% of tissue samples obtained in short-term survivors of GBM and in 15% of samples in the long-term survivors studied.¹
 - Inactivation of p14ARF was noted in 50% of primary and 75% of secondary GBMs.³

Bioinformatics methods are thus important in the generation and refinement of the project goals, and as will be seen later, are crucial in the development of the experimental techniques necessary to study the problem.

BIOINFORMATICS METHODS

We used RT-PCR assays to detect *Pokemon* expression in samples obtained in patients. The RT-PCR is still one of the most sensitive techniques for detection and quantitation

Abbreviations used in this paper: cDNA = complementary DNA; GBM = glioblastoma multiforme; mRNA = messenger RNA; NCBI = National Center for Biotechnology Information; RT-PCR = reverse transcription–polymerase chain reaction; UCSC = University of California, Santa Cruz.

of mRNA currently available. It will enable us to determine if *Pokemon* is expressed and, by measuring the expression of one of several housekeeping genes such as β -actin, *GAPDH*, or *cyclophilin*, determine if the level of expression is greater than in control tissues.

Briefly, the technique of RT-PCR begins with the isolation of RNA from the tissue samples. The mRNA is then reverse transcribed to produce cDNA. Once the mRNA has been reverse transcribed, it is necessary to use primers that are specific to the gene product of interest, which is *Pokemon* in this instance. The PCR is then used to amplify that specific gene product. The design of the specific primers needed to amplify the gene product of interest requires a bioinformatics approach, as outlined in the following section.

The first step was to determine the mRNA sequence for *Pokemon*. Using the NCBI, we searched for *ZBTB7A* in the Entrez Gene database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>) and selected "ZBTB7A: zinc finger and BTB domain containing 7A [*Homo sapiens*]." The Entrez Gene database (formerly LocusLink) serves to organize information about genes, serving as a major node in the nexus of genomic map, sequence, expression, protein structure, function, and homology data (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=handbook.TOC&depth=2> [NCBI Handbook]).

After finding the gene of interest, we scrolled down to the NCBI Reference Sequence (termed RefSeq on the site) and found the link for the corresponding mRNA. The Reference Sequence provides one accession number for a gene or protein that corresponds to the most stable, agreed-upon version of the sequence.⁴ Reference Sequence accession numbers for mRNA are NM_##### and for protein they are NP_#####. After one selects the mRNA accession number, the actual mRNA sequence from the GenBank database will be displayed. In this case, the accession number for *Pokemon* is NM_015898.

This sequence can then be copied into any of a number of primer design programs available on the Web. We chose to use Primer3 (found at http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).⁵ The mRNA sequence was pasted into the input box and the default parameters for primer design were accepted. The primers with the highest score were accepted.

In some instances, it is advantageous to select primers that span one or more introns. For example, if there is contaminating genomic DNA in the sample, the amplified DNA will contain the intron and will be much larger than expected. The amplified cDNA from the mRNA, which lacks the introns, will be the expected size. The intron-exon boundaries can be determined by doing a pairwise alignment of the gene and the cDNA with the Basic Local Alignment Search Tool (BLAST), which is located on the NCBI website (<http://www.ncbi.nlm.nih.gov/BLAST/>). As an alternative, we used the genome browser at the UCSC, with which we conducted a BLAST-Like Alignment Tool (BLAT) search by using both the complete *Pokemon* cDNA compared with genomic DNA (*ZBTB7A*) and the sequence amplified by the primers, and then we viewed a detailed alignment of the query sequence against genomic DNA (<http://genome.ucsc.edu/cgi-bin/hgBlat>). When it was determined that the initial primers accepted did not span an intron, we returned to Primer3 and designed primers by

specifying a region of the mRNA that corresponds to an intron-exon junction.

It is also desirable to determine whether the primer pair that has been selected will amplify other sequences in the genome (a potential source of confusion). Using the UCSC In-Silico PCR tool (<http://genome.ucsc.edu/cgi-bin/hgPcr?command=start>), we input both forward and reverse primers and searched the human genome to find possible amplification targets.

Using the available resources as outlined earlier, it is possible to design specific primers tailored to individual laboratory needs. In many cases, however, it is not necessary or desirable to custom-design primers but to use a published primer sequence, such as those for the primers that were designed for real-time PCR by Maeda, et al.,² for *Pokemon*. The resources listed earlier can be used to evaluate published primers or primers generated by electronic PCR on the NCBI website.

CONCLUSIONS

The application of mathematics, computers, and database technology to biology defines the field of bioinformatics, whereas the access to and analysis of the information generated and stored defines the method of bioinformatics. In this study, we found the following online sources to be of particular value: NCBI public databases, the genome browser at UCSC, and Primer3 (provided by the Whitehead Institute for Biomedical Research).

Using a variety of online databases, we were able to develop our research project: characterization of *Pokemon* expression in malignant glioma. We were also able to obtain the primer sequence information necessary for RT-PCR analysis of *Pokemon* expression. With our project as but one example, the promise of bioinformatics and the Human Genome Project has been fulfilled; stored and accessible genomic information has been translated into research of clinical relevance.

References

- Burton EC, Lamborn KR, Feuerstein BG, et al: Genetic aberrations defined by comparative genomic hybridization distinguish long-term from typical survivors of glioblastoma. **Cancer Res** 62:6205–6210, 2002
- Maeda T, Hobbs RM, Merghoub T, et al: Role of the proto-oncogene *Pokemon* in cellular transformation and ARF repression. **Nature** 433:278–285, 2005
- Nakamura M, Watanabe T, Klangbyu U, et al: p14ARF deletion and methylation in genetic pathways to glioblastoma. **Brain Pathol** 11:159–168, 2001
- Pevsner J: **Bioinformatics and Functional Genomics**. Hoboken, NJ: Wiley & Sons Inc, 2003, pp 29
- Rozen S, Skaletsky H: Primer3 on the WWW for general users and for biologist programmers, in Krawetz S, Misener S (eds): **Bioinformatics Methods and Protocols: Methods in Molecular Biology**. Totowa, NJ: Humana Press, 2000, pp 365–386
- Wiltshire RN, Herndon JE II, Lloyd A, et al: Comparative genomic hybridization analysis of astrocytomas: prognostic and diagnostic implications. **J Mol Diagn** 6:166–179, 2004

Manuscript received August 15, 2005.

Accepted in final form September 15, 2005.

Address reprint requests to: Richard A. Rovin, M.D., 580 West College Avenue, Marquette, Michigan 49855. email: rarovin@penmed.com.