Platform Transition:
New Technologies and Opportunities

Michael J. Kurtz & the ADS Team
mkurtz@cfa.harvard.edu

ADS Users Group Meeting - 11/3/2017

Change

- 1992 - The first Digital Libraries were just being started, conceived, ADS first public demonstration
- 2017 - The CFA Library now contains NO BOOKS
- 2042 - ?????

ADS Data

- A bibliographic database of 12M scientific publications in Astronomy and Physics (metadata)
- A full-text archive of 4.8M documents (XML, PDF, Latex, OCR text)
- A citation graph with over 6M nodes and 100M edges
- (Anonymous) usage data for 50k regular users

Machine Learning, Computational Linguistics

- Conditional Random Fields
- Latent Dirichlet Allocation (FS; O,T&S 1957, Topic Models)
- Word2Vec
- Various Ad Hoc

The new way: Machine Learning (CRF)

The AAPS Journal, Vol. 13, No. 2, June 2011 (© 2011)
DOI: 10.208/s12248-011-9280-2

Research Article

Evaluation of α 2-Integrin Expression as a Biomarker for Tumor Growth Inhibition for the Investigational Integrin Inhibitor E7820 in Preclinical and Clinical Studies

Ron J. Keizer,[1,6] Y. Funahashi,[2] T. Semba,[2] J. Wanders,[4] J. H. Beijnen,[1,5] J. H. M. Schellens[4,5] and A. D. R. Huitema[3]

Received 17 August 2010; accepted 4 February 2011; published online 9 March 2011

Abstract. E7820 is an orally active inhibitor of α2-integrin mRNA expression, currently tested in phases I and II. We aimed to evaluate what levels of inhibition in α2-integrin expression are needed to achieve tumor stasis in mice, and to compare this to the level of inhibition achieved in humans. Tumor growth inhibition was measured in mice bearing a pancreatic KP-1 tumor, dosed at 12.5–200 mg/kg/day for 21 days. In the phase I study, E7820 was administered daily for 28 days over a range of 0–200 mg, followed by a 7-day washout period. PK-PD models were developed in NONMEM. α2-Integrin expression measured on platelets, corresponding to tumor stasis at t=21 in 50% and 90% of the mice ($I_{50,ss}$, $I_{90,ss}$) were calculated. It was evaluated if these levels of inhibition could be achieved in patients at tolerable doses. One hundred nineteen α2-integrin measurements and 210 tumor size measurements were available from mice. The relationship between PK and α2-integrin expression was modeled using an indirect-effect model, subsequently linked to an exponential tumor growth model. $I_{50,ss}$ and $I_{90,ss}$ were 14.7% (RSE 7%) and 17.9% (RSE 8%). Four hundred sixty two α2-integrin measurements were available from 29 patients. Using the schedule of 100 mg qd (MTD), α2-integrin expression was inhibited more strongly than the $I_{50,ss}$ and $I_{90,ss}$ in greater than 95% and greater than 50% of patients, respectively. Moderate inhibition of α2-integrin expression corresponded to tumor stasis in mice, and similar levels could be reached in patients with the dose level of 100 mg qd.

KEY WORDS: biomarker; E7820; modeling and simulation; oncology; pharmacodynamics.

INTRODUCTION

The investigational anti-cancer drug E7820 is an orally active angiogenesis inhibitor, and has shown anti-tumor activity in several tumor models in mice, the effect of which was mediated through the inhibition of expression of α2-integrin (1,2). Integrins are receptors that mediate attachment between cells, and between cells and the extracellular matrix (3). They also play an important role in cell signaling. Many types of integrins have been identified, multiple types may be expressed on the cell surface simultaneously. Integrins are heterodimers, consisting of α (alpha) and β (beta) subunits. In mammals, 18 α and 8 β subunits have been characterized, with varying functions related to cell attachment and cell signaling. It has been shown in preclinical experiments that suppression of integrin α2 by E7820 played a crucial role in inhibition of endothelium tube formation (1,2). E7820 was evaluated in a phase I dose escalation study in patients with malignant solid tumors or lymphomas, the clinical results of which have been reported previously (4) and for which a population PK model was presented before (5). Since α2-integrin is expressed both on tumor cells and on platelets, it has been hypothesized that α2-integrin expression on platelets may be an easily evaluable biomarker for tumor growth inhibition in response to treatment with E7820 (1,2).

The inhibition of α2-integrin measured on platelets provides a measure of pharmacological target modulation, and therefore would theoretically provide a better predictor of activity than measures of E7820 plasma exposure. α2-Integrin is therefore currently being evaluated in phase I studies as possible biomarker (4,6). In this article, we present a modeling and simulation analysis that integrates data from preclinical experiments and a phase I clinical trial to evaluate the expected efficacy of the dose levels that were studies in phase I. Similar analyses have been presented earlier for the development of everolimus (7) and gefitinib (8). The aim of

[1] Department of Pharmacy and Pharmacology, The Netherlands Cancer Institute/Slotervaart Hospital, Louwesweg 6, 1066 EC, Amsterdam, The Netherlands.
[2] Research Laboratories Eisai Co., Ltd., Ibaraki, Japan.
[3] Eisai Co., Ltd. Hatfield, Hertfordshire, UK.
[4] Division of Clinical Pharmacology, Department of Medical Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
[5] Division of Drug Toxicology, Section of Biomedical Analysis, Departments of Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands.
[6] To whom correspondence should be addressed. (e-mail: ron.keizer@bls.nl)

The AAPS Journal, Vol. 13, No. 2, June 2011 (© 2011)
DOI: 10.208/s12248-011-9280-2

Research Article

Evaluation of α 2-Integrin Expression as a Biomarker for Tumor Growth Inhibition for the Investigational Integrin Inhibitor E7820 in Preclinical and Clinical Studies

Ron J. Keizer, Y. Funahashi, T. Semba, J. Wanders, J. H. Beijnen, J. H. M. Schellens and A. D. R. Huitema

Received 17 August 2010; accepted 4 February 2011; published online 9 March 2011

Abstract. E7820 is an orally active inhibitor of α2-integrin mRNA expression, currently tested in phases I and II. We aimed to evaluate what levels of inhibition in α2-integrin expression are needed to achieve tumor stasis in mice, and to compare this to the level of inhibition achieved in humans. Tumor growth inhibition was measured in mice bearing a pancreatic KP-1 tumor, dosed at 12.5–200 mg/kg/day for 21 days. In the phase I study, E7820 was administered daily for 28 days over a range of 0–200 mg, followed by a 7-day washout period. PK-PD models were developed in NONMEM. α2-Integrin expression measured on platelets, corresponding to tumor stasis at t=21 in 50% and 90% of the mice ($I_{50,ss}$, $I_{90,ss}$) were calculated. It was evaluated if these levels of inhibition could be achieved in patients at tolerable doses. One hundred nineteen α2-integrin measurements and 210 tumor size measurements were available from mice. The relationship between PK and α2-integrin expression was modeled using an indirect-effect model, subsequently linked to an exponential tumor growth model. $I_{50,ss}$ and $I_{90,ss}$ were 14.7% (RSE 7%) and 17.9% (RSE 8%). Four hundred sixty two α2-integrin measurements were available from 29 patients. Using the schedule of 100 mg qd (MTD), α2-integrin expression was inhibited more strongly than the $I_{50,ss}$ and $I_{90,ss}$ in greater than 95% and greater than 50% of patients, respectively. Moderate inhibition of α2-integrin expression corresponded to tumor stasis in mice, and similar levels could be reached in patients with the dose level of 100 mg qd.

KEY WORDS: biomarker; E7820; modeling and simulation; oncology; pharmacodynamics.

INTRODUCTION

The investigational anti-cancer drug E7820 is an orally active angiogenesis inhibitor, and has shown anti-tumor activity in several tumor models in mice, the effect of which was mediated through the inhibition of expression of α2-integrin (1,2). Integrins are receptors that mediate attachment between cells, and between cells and the extracellular matrix (3). They also play an important role in cell signaling. Many types of integrins have been identified, multiple types may be expressed on the cell surface simultaneously. Integrins are heterodimers, consisting of α (alpha) and β (beta) subunits. In mammals, 18 α and 8 β subunits have been characterized, with varying functions related to cell attachment and cell signaling. It has been shown in preclinical experiments that suppression of integrin α2 by E7820 played a crucial role in inhibition of endothelium tube formation (1,2). E7820 was evaluated in a phase I dose escalation study in patients with malignant solid tumors or lymphomas, the clinical results of which have been reported previously (4) and for which a population PK model was presented before (5). Since α2-integrin is expressed both on tumor cells and on platelets, it has been hypothesized that α2-integrin expression on platelets may be an easily evaluable biomarker for tumor growth inhibition in response to treatment with E7820 (1,2).

The inhibition of α2-integrin measured on platelets provides a measure of pharmacological target modulation, and therefore would theoretically provide a better predictor of activity than measures of E7820 plasma exposure. α2-Integrin is therefore currently being evaluated in phase I studies as possible biomarker (4,6). In this article, we present a modeling and simulation analysis that integrates data from preclinical experiments and a phase I clinical trial to evaluate the expected efficacy of the dose levels that were studies in phase I. Similar analyses have been presented earlier for the development of everolimus (7) and gefitinib (8). The aim of

[1] Department of Pharmacy and Pharmacology, The Netherlands Cancer Institute/Slotervaart Hospital, Louwesweg 6, 1066 EC, Amsterdam, The Netherlands.
[2] Research Laboratories Eisai Co., Ltd., Ibaraki, Japan.
[3] Eisai Co., Ltd. Hatfield, Hertfordshire, UK.
[4] Division of Clinical Pharmacology, Department of Medical Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
[5] Division of Drug Toxicology, Section of Biomedical Analysis, Department of Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands.
[6] To whom correspondence should be addressed. (e-mail: ron.keizer@bls.nl)

230

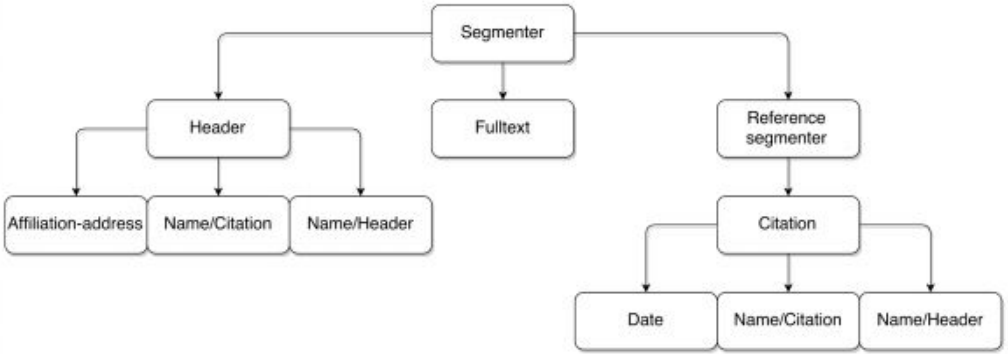
aaps

abstract

Courtesy: P. Lopez

Grobid (http://cloud.science-miner.com/grobid/)

- 11 CRF models (2 for patents)
- Full text processing uses 9 models, 55 final labels, 14 intermediary labels
- More than 10 000 labelled examples



science-miner

Courtesy: P. Lopez

Spark MLlib

Download | Libraries | Documentation | Examples | Community | Developers

MLlib is Apache Spark's scalable machine learning library.

Ease of Use

Usable in Java, Scala, Python, and R.

MLlib fits into Spark's APIs and interoperates with NumPy in Python (as of Spark 0.9) and R libraries (as of Spark 1.5). You can use any Hadoop data source (e.g. HDFS, HBase, or local files), making it easy to plug into Hadoop workflows.

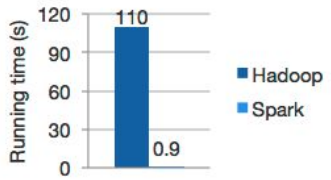
```
data = spark.read.format("libsvm")\
  .load("hdfs://...")

model = KMeans(k=10).fit(data)
```

Calling MLlib in Python

Performance

High-quality algorithms, 100x faster than MapReduce.

Spark excels at iterative computation, enabling MLlib to run fast. At the same time, we care about algorithmic performance: MLlib contains high-quality algorithms that leverage iteration, and can yield better results than the one-pass approximations sometimes used on MapReduce.



Logistic regression in Hadoop and Spark

Graph/Network Analysis

- Very complex multi-partite network
- Clustering/Community Detection
- Information Flow, Critical Paths

Apache Spark GraphX

Download | Libraries | Documentation | Examples | Community | Developers

GraphX is Apache Spark's API for graphs and graph-parallel computation.

Flexibility

Seamlessly work with both graphs and collections.

GraphX unifies ETL, exploratory analysis, and iterative graph computation within a single system. You can view the same data as both graphs and collections, transform and join graphs with RDDs efficiently, and write custom iterative graph algorithms using the Pregel API.

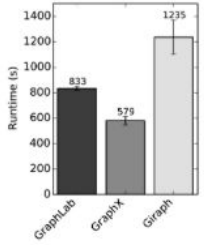
```
graph = Graph(vertices, edges)
messages = spark.textFile("hdfs://...")
graph2 = graph.joinVertices(messages) {
  (id, vertex, msg) => ...
}
```

Using GraphX in Scala

Speed

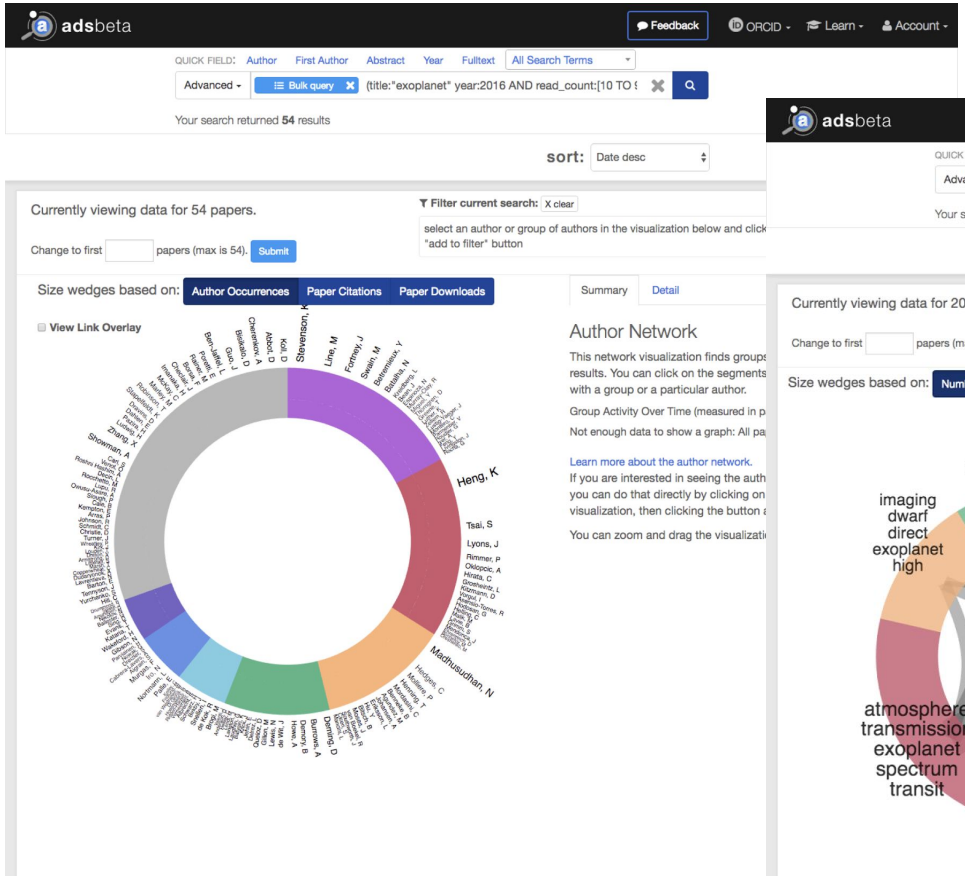
Comparable performance to the fastest specialized graph processing systems.

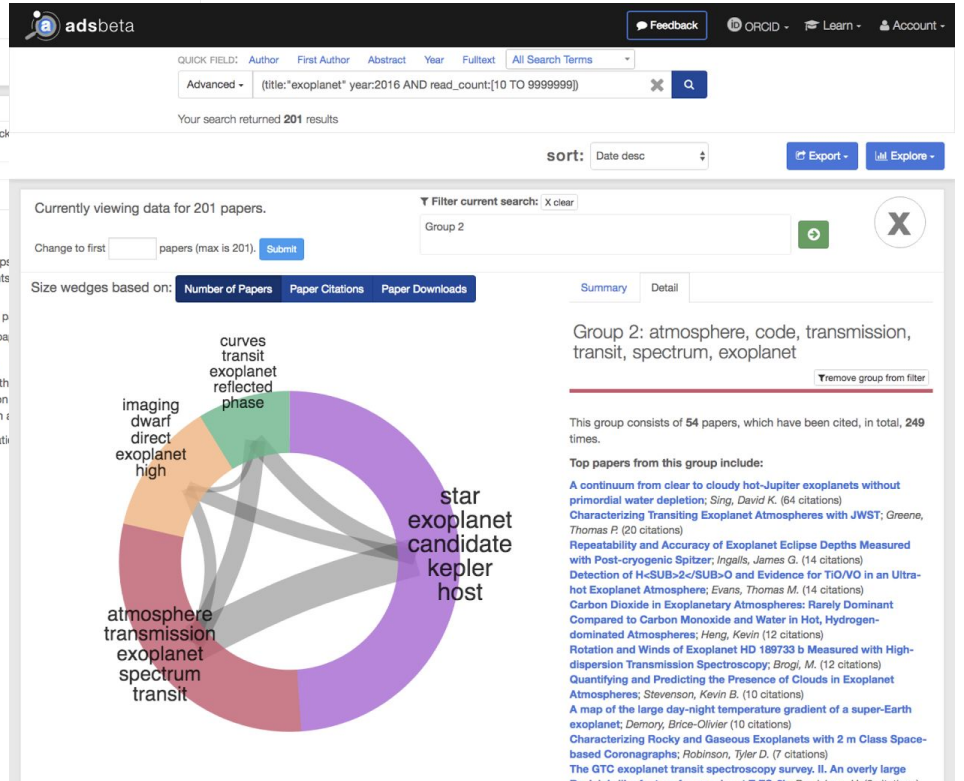
GraphX competes on performance with the fastest graph systems while retaining Spark's flexibility, fault tolerance, and ease of use.

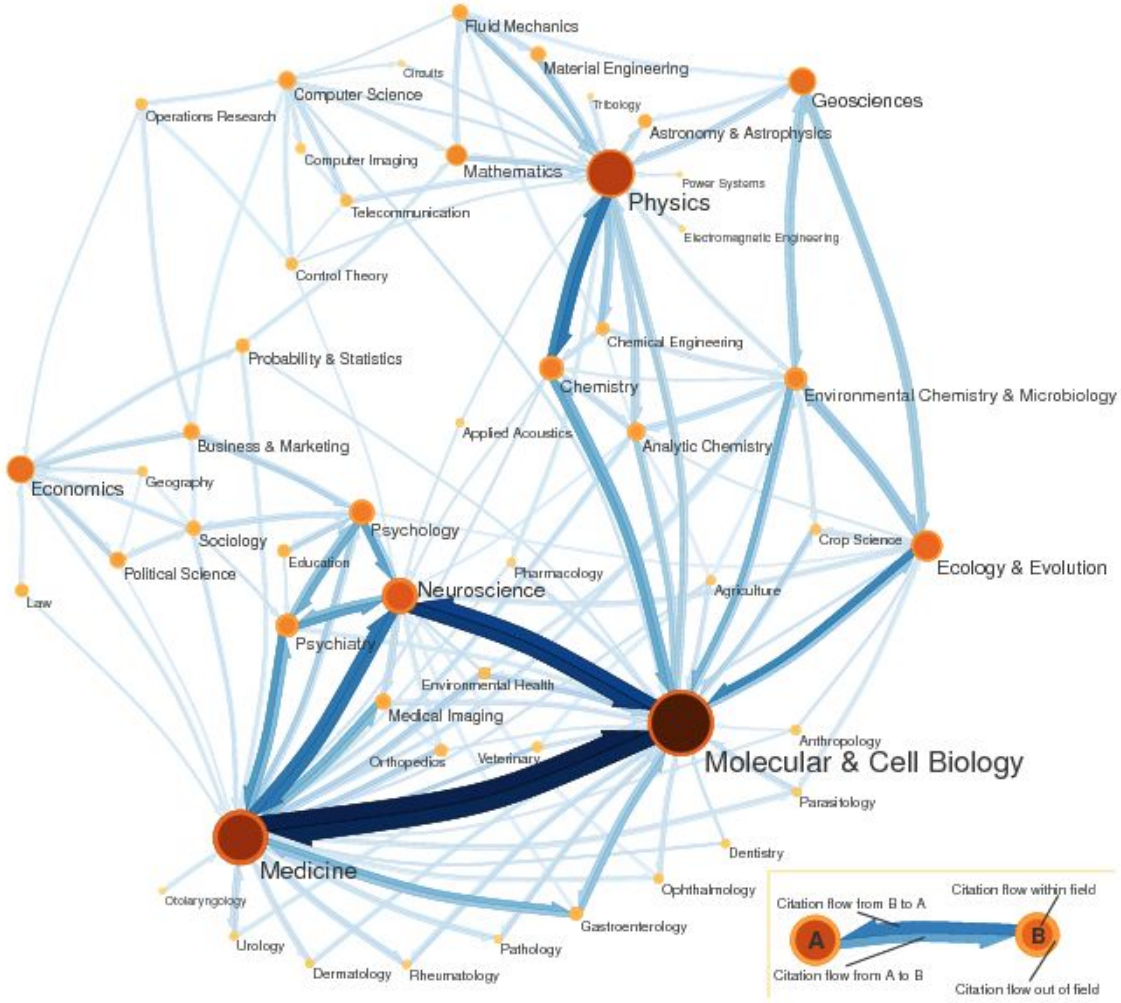


End-to-end PageRank performance (20 iterations, 3.7B edges)

Network Visualizations






Fluid Mechanics
Material Engineering
Circuits
Computer Science
Geosciences
Operations Research
Tribology
Astronomy & Astrophysics
Computer Imaging
Mathematics
Power Systems
Physics
Telecommunication
Electromagnetic Engineering
Control Theory
Chemical Engineering
Probability & Statistics
Chemistry
Environmental Chemistry & Microbiology
Applied Acoustics
Business & Marketing
Analytic Chemistry
Economics
Geography
Psychology
Crop Science
Sociology
Education
Political Science
Pharmacology
Ecology & Evolution
Neuroscience
Agriculture
Law
Psychiatry
Environmental Health
Medical Imaging
Anthropology
Orthopedics
Veterinary
Molecular & Cell Biology
Parasitology
Dentistry
Medicine
Ophthalmology
Otolaryngology
Gastroenterology
Urology
Pathology
Dermatology
Rheumatology
Citation flow from B to A
Citation flow within field
A
B
Citation flow from A to B
Citation flow out of field

Human-Computer Interface

- Visualizations
- User Interfaces

Paper analytics

How it's done: paper to electronic record

The Astronomical Journal, 124:266–293, 2002 July

DETAILED STRUCTURAL DECOMPOSITION OF GALAXY IMAGES[1]

Chien Y. Peng,[2] Luis C. Ho,[3] Chris D. Impey,[2] and Hans-Walter Rix[4]

Received 2001 August 16; accepted 2002 March 27

ABSTRACT

We present a two-dimensional fitting algorithm (GALFIT) designed to extract structural components from galaxy images, with emphasis on closely modeling light profiles of spatially well-resolved, nearby galaxies observed with the *Hubble Space Telescope*. Our algorithm improves on previous techniques in two areas: by being able to simultaneously fit a galaxy with an arbitrary number of components and with optimization in computation speed, suited for working on large galaxy images. We use two-dimensional models such as the "Nuker" law, the Sérsic (de Vaucouleurs) profile, an exponential disk, and Gaussian or Moffat functions. The azimuthal shapes are generalized ellipses that can fit disky and boxy components. Some potential applications of our program include: standard modeling of global galaxy profiles; extracting bars, stellar disks, double nuclei, and compact nuclear sources; and measuring absolute dust extinction or surface brightness fluctuations after removing the galaxy model. When examined in detail, we find that even simple looking galaxies generally require at least three components to be modeled accurately, rather than the one or two components more often employed. Many galaxies with complex isophotes, ellipticity changes, and position angle twists can be modeled accurately in two dimensions. We illustrate this by way of 11 case studies, which include regular and barred spiral galaxies, highly disky lenticular galaxies, and elliptical galaxies displaying various levels of complexities. A useful extension of this algorithm is to accurately extract nuclear point sources in galaxies. We compare two-dimensional and one-dimensional extraction techniques on simulated images of galaxies having nuclear slopes with different degrees of cuspiness, and we then illustrate the application of the program to several examples of nearby galaxies with weak nuclei.

Key words: galaxies: bulges — galaxies: fundamental parameters — galaxies: nuclei — galaxies: structure — techniques: image processing — techniques: photometric

1. INTRODUCTION

Galaxies span a wide range of morphology and luminosity, and a very useful way to quantify them is to fit their light distribution with parametric functions. The de Vaucouleurs $R^{1/4}$ and exponential disk functions became standard functions to use after de Vaucouleurs (1948) showed many elliptical galaxies to have $R^{1/4}$ light distributions, while Freeman (1970) found later-type galaxies to be well described by a de Vaucouleurs bulge plus an exponential disk. Since then the empirical techniques of galaxy fitting and decomposition have led to a number of notable advances in understanding galaxy formation and evolution. These include investigations into the Tully-Fisher relationship (Tully & Fisher 1977), the fundamental plane of spheroids (Faber et al. 1987; Dressler et al. 1987; Djorgovski & Davis 1987; Bender, Burstein, & Faber 1992), the morphological transformation of galaxies in cluster environments (e.g., Dressler 1980; van Dokkum & Franx 2001), the bimodality of galaxy nuclear cusps (Lauer et al. 1995; Faber et al. 1997) and its implications for the formation of massive black holes (Ravindranath, Ho, & Filippenko 2002), and the cosmic evolution of galaxy morphology (e.g., Lilly et al. 1998; Marleau & Simard 1998).

There are two general types of galaxy fitting: one-dimensional fitting of surface brightness profiles (e.g., Kormendy 1977; Burstein 1979; Boroson 1981; Kent 1985; Baggett, Baggett, & Anderson 1998) and two-dimensional fitting of galaxy images (e.g., Shaw & Gilmore 1989; Byun & Freeman 1995; de Jong 1996; Simard 1998; Wadadekar, Robbason, & Kembhavi 1999; Khosroshahi, Wadadekar, & Kembhavi 2000), each with its own tradeoffs and benefits.

In one dimension an important consideration is how to first obtain a radial surface brightness profile from a two-dimensional image, for which there is no universally agreed upon procedure. A common practice is to use isophote fitting, which is a powerful technique when performed on well-resolved galaxies because it averages over elliptical annuli to increase the signal-to-noise ratio (S/N) at a given radius. However, as many galaxies have isophote twists and changing ellipticity as a function of radius, the galaxy profile is extracted along a radial arc that is ill defined. An alternative approach is to use a direct one-dimensional slice across an image. Burstein (1979) argues that only cuts along the major axis should be used in bulge-to-disk (B/D) decompositions. Meanwhile, Ferrarese et al. (1994) point out that galaxies with power-law central profiles may have different profiles along the major and minor axis.

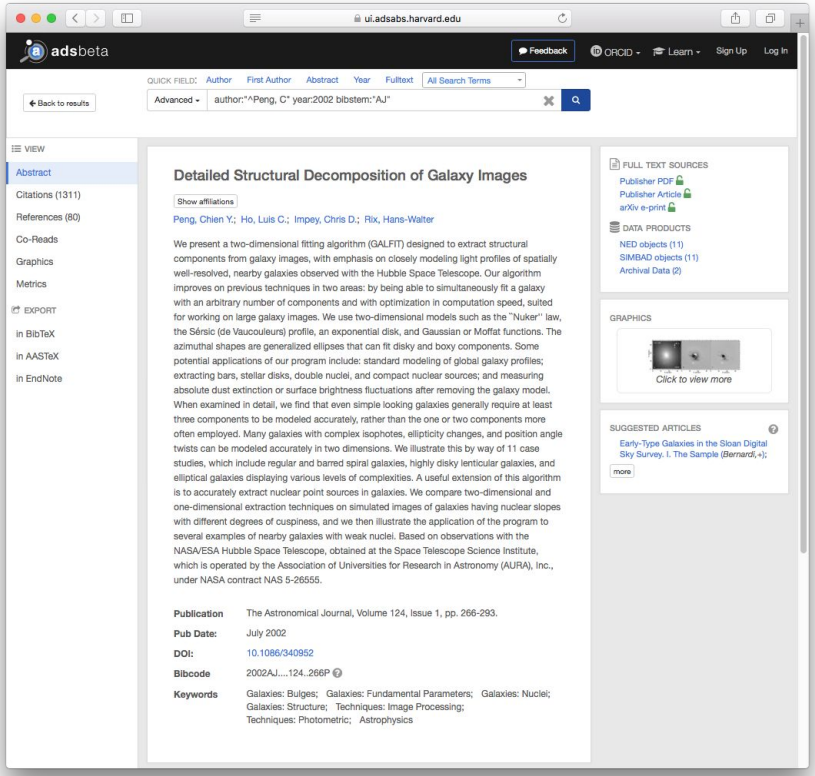
Fitting profiles in one dimension is frequently used because it suffices for certain goals and is simple to implement. But many studies now resort to two-dimensional techniques. For B/D decompositions a number of authors (e.g., Byun & Freeman 1995; Wadadekar et al. 1999) have

[1] Based on observations with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy (AURA), Inc., under NASA contract NAS 5-26555.

[2] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721; cyp, cimpey@as.arizona.edu.

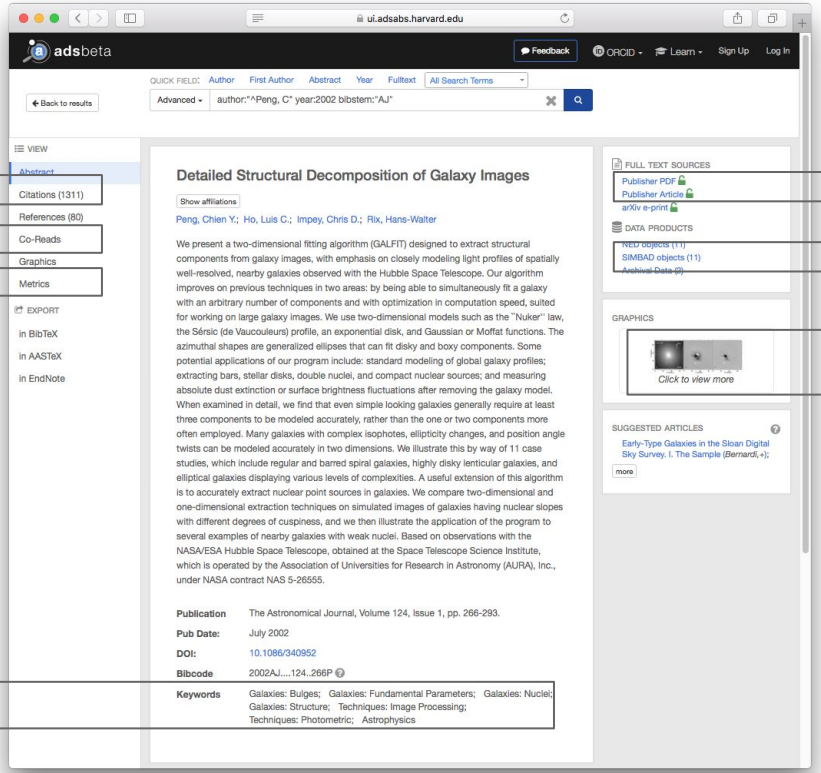
[3] The Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101; lho@ociw.edu.

[4] Max-Planck-Institut für Astronomie, Königstuhl 17, Heidelberg, D-69117, Germany.



Mining, collecting and connecting

Citations/References
Co-Reads (usage)
Metrics



Topics (keywords)

Links to Full Text

Links to Data

Figures

Future

- When Computing and Storage are *ANOTHER* million times more powerful, what will ADS be like?
- ADS is a Human-Machine collaboration
- The New-Gutenbergian Revolution is not yet finished
- ADS is where the state-of-the-art becomes the state-of-the-practice

For More Information

- ADS Bumblebee:
  https://ui.adsabs.harvard.edu
- ADS API:
  https://github.com/adsabs/adsabs-dev-api
- ADS help and support:
  http://adsabs.github.io/help/ adshelp@cfa.harvard.edu
- ADS news:
  http://adsabs.github.io/blog/ @adsabs
- ADS users group:
  http://adsabs.harvard.edu/adsug.html