

7. ГЕНОМНАЯ МЕДИЦИНА И ГРИД-КОМПЬЮТИНГ

Полная реализация концепции *геномной медицины*, в которой для поддержки здравоохранения используются геномика и протеомика, требует интеграции знаний из областей, традиционно существующих отдельно, а именно из биологии и медицины. Чтобы эффективно использовать все богатство информации, которой располагают исследовательские центры и медицинские учреждения, должны быть созданы компьютерные методы и инструментальные программные средства, которые позволят навести мосты между медицинской информатикой и биоинформатикой.

При таком подходе можно было бы использовать все уровни информации – от молекулярного до уровня популяции через клетки, ткани, органы и пациента – и наиболее подходящие приемы и методы. Некоторые из этих методов могли бы заимствоваться из биоинформатики, другие – из медицинской информатики и даже из общественного здравоохранения или эпидемиологической информатики (см. Таблица 1).

7.1. РАЗРАБОТКИ В ГЕНОМИКЕ, ВЛИЯЮЩИЕ НА МЕДИЦИНСКОЕ ОБСЛУЖИВАНИЕ

Завершенный проект “Геном человека” (HGP - Human Genome Project) станет для медицины источником новых знаний о взаимоотношениях между структурой генов человека, факторами окружающей среды и физиопатологическими процессами [1]. Постгеномная эра открывает возможности для изучения всех генов, всех белков и большого числа мутаций в клетках человека, а это открывает путь к недоступным до настоящего времени методам исследований, которые должны привести к пониманию молекулярной основы сложных заболеваний и, тем самым, способствовать появлению новых решений в диагностике и терапии [2].

Геномная медицина повлияет на разные области медицинского обслуживания:

Клиническая диагностика: Новые высокопроизводительные устройства, применяемые при клиническом обследовании (биочипы), позволяют проводить одновременный мониторинг большого числа параметров, которые могут использоваться как диагностические маркеры. Генетический анализ используется для идентификации отдельных людей с подозрением на заболевание и для подтверждения подозреваемой мутации у человека или семьи до того, как появятся симптомы заболевания [4]. Новые маркеры, представляющие интерес для мониторинга пациента, предложит также и протеомика [5].

Переклассификация болезни: Сравнение различных профилей экспрессии генов для здоровых клеток и клеток из тканей, пораженных болезнью, позволит в некоторых случаях идентифицировать различные формы молекул и по-новому классифицировать болезни, что приведет к уточнению диагноза.

Фармакогенетика и фармакогеномика: За последние несколько лет были разработаны успешные технологические методы для изучения и применения индивидуальных изменений на молекулярном уровне. Новые технологии, которые способствуют нашему пониманию роли генов в возникновении болезней, обеспечивают промышленности реальные возможности для производства более сильных лекарственных препаратов,

безопасных лекарств и улучшенных вакцин (*фармакогеномика*) [6].

Генетическая эпидемиология и общественное здравоохранение: Использование новых технологий в области генетической информации даст возможность проводить экономически эффективный скрининг (генетические тесты) населения [7]. Чтобы передать знания, полученные геномикой, общественному здравоохранению и эпидемиологии, важно будет проводить работы в области ассоциативной генетики, изучать взаимоотношения генотип-фенотип для населения и создавать программы распространения генетической информации и обучения медицинского персонала.

Исследования, проводящиеся в настоящее время в геномной медицине, порождают огромный объем данных, для обработки которых требуются глобальные распределенные ресурсы и развитые инструментальные средства [8].

7.2. КОНВЕРГЕНЦИЯ БИО - И МЕДИЦИНСКОЙ ИНФОРМАТИКИ

Термин 'биомедицинская информатика' все чаще используется на конференциях и в статьях, обозначая то пространство, где встречаются и взаимодействуют медицинская информатика и биоинформатика.

Современная биоинформатика пользуется банками данных в интернете, откуда все научное сообщество может получать нужную информацию. Однако, имеющимся в настоящее время средствам информатики не хватает необходимых методов и средств для связывания генетической и клинической информации и для того, чтобы медицинские приложения использовали существующие генетические базы данных [9].

Необходимы средства управления данными, которые позволят преобразовать огромный объем информации, получаемый генетиками и молекулярными биологами в своих лабораториях, в данные, которые смогут использовать врачи и медицинский персонал. Главная задача на сегодняшний день – найти такие технологии, которые позволят трансформировать достижения в биомедицине в разделяемые знания, способствующие правильному диагностированию и лечению.

Хотя в настоящее время трудно предсказать, какие медицинские проблемы может породить мутация одного гена или белка, и как можно будет использовать это знание в клинической практике, уже ясно, что гены взаимодействуют с многими другими генами и с окружающей средой. Только объединение изучения взаимодействия генов у человека и животных с эпидемиологическими исследованиями многих различных популяций сможет раскрыть сложный механизм развития генетических заболеваний.

Достижения в понимании генетического кода, генных продуктов и функций способствуют разъяснению механизмов развития болезней. Целостный взгляд на здоровье отдельного человека формируется на базе различных источников знаний в сочетании с клинической и генетической информацией. Биомедицинская информация, доступная исследователям и практикам, содержит данные о пациенте, о геноме и последовательностях, о структуре и последовательностях белков, о мутациях, генетических болезнях, генетических тестах, о терминологии и системах кодирования, о рекомендациях пациентам и многое другое.

Навигация между *фенотипом* и *генотипом* в клинических условиях означает, что генетическая оценка должна быть интегрирована в обследование пациентов. Для этого

требуется разработка и реализация компьютерных методов и инструментальных средств для бесшовного объединения биомедицинских данных. Интеграция биомедицинских знаний порождает новый круг проблем, среди которых особое внимание должно быть уделено следующим:

- в вебе имеется много различных источников информации; нужно смоделировать релевантную информацию, найти её и выбрать;
- интеграция данных затрудняется тем, что базы данных могут иметь разный формат и разную семантику. Кроме того, информация в вебе часто доступна только через веб-интерфейсы, и компьютерным приложениям не так легко до нее добраться;
- способы кодирования и терминология не унифицированы, например, далеки от унификации названия генов;
- медицинские системы не готовы к работе с генетической информацией;
- необходимость соблюдения прав интеллектуальной собственности, секретности и конфиденциальности информации и защиты права собственности на данные может затруднить обмен информацией;
- результаты исследований часто публикуются на естественных языках (научная библиография), поэтому нужны средства для представления полученных знаний в машинно-читаемом виде;
- объем данных, подлежащих обработке, огромен, поэтому требуются высокопроизводительные хранилища данных, вычислительные мощности и сетевые инфраструктуры, гарантирующие эффективное управление данными и их использование.

Уровень медицинской информации	Классические приложения в медицинской информатике	Новые геномные данные и информация	Новые приложения в медицинской информатике
Популяция	<ul style="list-style-type: none"> - Общественное здравоохранение & эпидемиологические базы данных - Оценка технологии, исследование результатов 	<ul style="list-style-type: none"> - Геномная эпидемиология - Генетический скрининг 	<ul style="list-style-type: none"> - Базы данных для геномной эпидемиологии и сеть (CDC-HuGeNet)
Болезни	<ul style="list-style-type: none"> - Системы классификации болезней - Руководства по использованию компьютеров в клинической практике (CCPGs) 	<ul style="list-style-type: none"> - Новая классификация болезней по их молекулярной основе - Принятие решений на генетической основе - Клинические исследования в фармакогенетике 	<ul style="list-style-type: none"> - Инструментальные средства для поддержки принятия решений - Молекулярная классификация болезней - CCPG, содержащие генетические тесты и последующее лечение

	- Информационные системы для клинических исследований		на базе генетических данных - Базы данных для фармакогенетики
Пациент	- Электронная история болезни пациента (CPHR)	- Индивидуальные генетические профили (SNP, мутации)	- Генетические данные в CPHR
Ткани, органы	- Системы для патологолабораторий, обработка медицинских изображений	- Физиологическая геномика - Генетические сети	- Банки данных для опухолей - Модели болезней
Клетки	- Медицинские изображения в цитогенетике, гистология - Информационные системы для лабораторий микробиологии	- Профили экспрессии генов - Протеомика	- Молекулярные изображения - Информационные системы для фармакогеномики (R&D для создания лекарств)
Молекулярный уровень	- Биохимические и генетические тесты и информационные системы для лабораторий	- ДНК и последовательности белков - Микромолекулярные структуры	- Облегчение доступа к нужным базам данных по геномике для профессионалов здравоохранения

Таблица 1: Совместные усилия медицинской информатики и биоинформатики по расширению возможностей информатики для здравоохранения (см. [10])

7.3. СЕМАНТИЧЕСКАЯ ИНТЕГРАЦИЯ БИМЕДИЦИНСКИХ РЕСУРСОВ

Информационные биомедицинские ресурсы обычно не связаны друг с другом, хотя их содержимое семантически родственно. Сведение воедино таких знаний является сложной задачей, поскольку трудно автоматически устанавливать семантические связи.

Семантическая интеграция таких ресурсов стала бы одним из факторов, способствующих разработке новейших биомедицинских приложений для исследовательских центров, специализированных базовых лечебных учреждений и лабораторий, а также для центров здравоохранения, где применяются клинические методические указания, например, для больниц. Главными целями семантической интеграции информационных биомедицинских ресурсов являются:

- согласованный и понятный доступ к биологическим, биомедицинским, медицинским и клиническим ресурсам, к ресурсам биоинформатики, особенно к банкам данных (например, SwissProt, PDB - Protein Data Bank) и системам электронных историй болезни пациентов (EPR - Electronic Patient Record) [11];
- содействие в поиске и использовании семантических связей внутри источников данных и между такими источниками (например, последовательность белков в SwissProt родственна вторичной структуре белков в PDB, а трехмерная форма

белков в PDB может быть связана с лекарственными соединениями в базе данных лигандов).

Семантической интеграции биомедицинских ресурсов могут принести пользу существующие стандарты, такие, как управление появляющимися знаниями, методология и технология работы с моделями (системы извлечения информации и текстов из данных, системы управления документами и их содержимым), онтологии, реляционные базы данных, полуструктурированные базы данных, языки моделирования и управление метаданными. К основным службам, определяющим основу семантической интеграции, относятся:

- семантическое моделирование различных биомедицинских понятий и ресурсов с использованием онтологий (таких как GeneOntology [12]) и метаданных;
- семантическая аннотация биомедицинских ресурсов, делающая возможным непрерывный обмен знаниями между источниками данных и пользователями (исследователями, врачами, медиками и т.д.);
- поиск, просмотр и запрос биомедицинских ресурсов, осуществляемый как пользователями, так и компьютерными программами не по ключевым словам, а по семантическим понятиям;
- семантическое моделирование медицинской документации с использованием различных типов метаданных: зависимость от типа носителя документации, описание содержимого документа, классификация содержимого, состав документа, история документа, расположение документа.

Последние достижения грид-технологий соответствуют потребностям семантической интеграции. Новые возникающие грид-инфраструктуры включают в себя:

- веб-службы, которые осуществляют поиск, вызов и выполнение распределенных служб и могли бы использоваться для реализации некоторых базовых биомедицинских служб и приложений;
- базирующиеся на гриде СУБД и системы управления метаданными. Чтобы обеспечить надежное, эффективное и автоматическое управление источниками данных в среде грида, можно ввести новое понятие: грид-СУБД [13];
- поддержка кластеров виртуальных организаций через основные грид-службы, такие как обеспечение безопасности, и инструментальные средства и платформы для объединения организаций.

Семантическая интеграция включает в себя как моделирование, так и технологию. Моделирование способствует внедрению семантических служб высокого уровня и приложений, а технология может повысить производительность и эффективность при работе в распределенной среде и среде грида.

7.4. БИОМЕДИЦИНСКИЕ ГРИДЫ ДЛЯ МЕДИЦИНСКИХ ПРИЛОЖЕНИЙ

Геномная медицина нуждается в развитии многих областей информатики, куда входят численное моделирование, получение молекулярных изображений, широкомасштабный доступ к данным и т.д. [14]. К этим областям относятся и грид-технологии, которые могут внести вклад в развитие некоторых ключевых направлений благодаря тому, что (1) предоставляют мощные вычислительные возможности; (2) обеспечивают бесшовный доступ и интеграцию сложных и распределенных источников данных; (3) поддерживают создание объединенных виртуальных организаций, позволяющее улучшить взаимодействие на уровне человек-человек [15] [16].

Ожидаемый вклад грид-технологий в развитие геномной медицины:

1. **Вычислительная геномика и вычислительная протеомика** в идентификации генов и белков, автоматическое составление аннотаций и характеристик индивидуальных генетических изменений (например, виртуальные лаборатории генетической информации).
2. **Технологии хранения больших объемов данных о фенотипах, генотипах и протеотипах** в метареляционных базах данных.
3. **Поддержка проведения клинических исследований.**
4. **Обеспечение персонализированных медицинских услуг** при помощи генетического профилирования пациента, понимания его наследственности, проведения согласованных клинических наблюдений, эпидемиологических исследований и статистического анализа.
5. **Разработка моделей и численных имитаций клеток и болезней.** Связывание образцов экспрессии генов с моделями болезней для обнаружения путей развития болезни, определяемых клиническим состоянием пациента, его образом жизни, питанием и генетической предрасположенностью к заболеванию. Повсеместный доступ к полной истории болезни пациента, независимо от того, в каком центре была собрана информация о болезни и лечении пациента. Для изучения путей и процессов обмена веществ можно построить трехмерные модели (тела, клеток и т.д.), соединяющие анатомические и функциональные параметры и связывающие структурную информацию с информацией о совокупности клеток. Наличие соответствующих вычислительных ресурсов позволило бы постепенно формировать унифицированную абстракцию из последовательности генов, функций, патофизиологических процессов и клинических проявлений болезни. Эта функциональная модель могла бы стать источником справочной информации для преподавателей и профессионалов в области здравоохранения. Такие системы будут использоваться при оценке воздействия токсичных веществ или реакции клеток на лекарства (например, [17]).
6. **Предоставление программных средств для обучения врачей и улучшения управления биомедицинскими знаниями.** Большинство врачей имеет только элементарное представление о генетике и геномике. Программы для е-обучения могут иметь решающее значение для легкого и быстрого внедрения новых методов и новых разработок в повседневную практику и для знакомства с геномной медициной. Эти программы могли бы работать с разделяемыми ресурсами, такими как файлы данных и модели, и

использовать грид-технологии, например, средства интеграции и разделения ресурсов. Должны быть созданы порталы для е-здравоохранения, ориентированные на решение задач здравоохранения с помощью распределенных приложений.

7. Молекулярные изображения. Соединение технологий получения медицинских изображений и геномных подходов порождает новую область функциональных и молекулярных изображений. Эта область может пополнить диагностический арсенал путем визуализации *in vivo* процессов, происходящих в клетках и генах. Целью молекулярного изображения является количественное и бескровное изучение болезни на молекулярном уровне. Грид может обеспечить необходимые мощности для обработки данных в этой области.

8. Генетическая эпидемиология. Может быть проведено изучение влияния генетических факторов и факторов окружающей среды на развитие отдельных болезней в группах населения. Источники информации, необходимые для таких исследований, находятся в различных и удаленных местах. Грид-инфраструктуры могут облегчить бесшовный доступ ко всем этим информационным ресурсам.

9. Развитие фармакогеномики. Конструирование лекарств может быть коренным образом изменено, если применять не традиционный метод проб и ошибок, а новый обоснованный подход, использующий информацию о последовательностях генов и структуре белков. Для работы со сложным спектром источников информации, требующейся при таком подходе (лабораторные измерения, клинические показатели, генетические изменения у человека, химические соединения и направленность метаболизма) потребуется новое поколение моделей данных и их хранилищ. Грид предлагает услуги, которые помогут справиться с многообразием источников информации.

10. Разработка программных средств поддержки принятия решений в клиниках, объединение многочисленных источников релевантной информации (генетической, клинической и информации об окружающей среде). В рамках геномной медицины медики будут иметь доступ к биологической информации и смогут объединять её с данными, содержащимися в электронных историях болезни пациентов, или в информационных системах отделений больших больниц. Грид мог бы оказать помощь в интеграции всех данных, используемых для принятия решений, и предоставить вычислительные мощности, необходимые для работы сложных интерактивных систем в режиме реального времени.

11. Интеграция баз данных и знаний, относящихся к клинической практике и геномным исследованиям. Биомедицина – это наука, где ведутся совместные исследования, объединяющие знания и ресурсы коллективов, представляющих разные научные дисциплины. Часто для проведения исследований различные институты объединяются и создают виртуальную организацию. Партнерам, занятым в биомедицинских исследованиях, требуется вычислительная инфраструктура, которая может поддерживать такой вид объединения и совместного использования информационных систем, часто уже существующих, разнородных и децентрализованных. Кроме того, прогресс в науке зависит от способности создавать общие представления (онтологии, объединенные словари и т.д.) для моделирования и описания гетерогенной информации. Главная задача – адаптировать существующие системы или создавать новые,

которые сделают возможным обмен данными и их интеграцию. Грид, усиленный службами семантической интеграции, может помочь не только в разделении ресурсов, но и в интеграции генетических данных, полученных в функциональной и сравнительной (индивидуальной) геномике, в клинические информационные системы.

7.5. ПОТРЕБНОСТИ И КОНФИГУРАЦИИ БИМЕДИЦИНСКИХ ГРИДОВ

Путь к эффективной работе с данными на разных уровнях грида (запрос, представление, обмен, объединение и преобразование в полезное знание) открывает новая область исследований, известная как "*Интеллектуальный грид*" [19]. В частности, могут быть созданы службы Интеллектуального грида на базе онтологий и метаданных [20]. Используя онтологии, гриды могут осуществлять семантическое моделирование задач/потребностей пользователя, предоставлять услуги и источники данных для поддержки грид-служб высокого уровня и выполнять динамический поиск и композицию служб. Кроме того, используя средства извлечения информации из данных и управления знаниями, можно было бы создавать новейшие службы, базирующиеся на семантике хранимых данных. *Семантический грид* сосредоточивается на систематическом обращении к метаданным и онтологиям для описания грид-ресурсов, для улучшения и автоматизации поиска служб и согласования их функций, для композиции приложений, извлечения информации и обнаружения знаний [21]. *Гриды знаний* [22] предлагают средства высокого уровня для распределенного поиска и извлечения знаний из репозитория данных, доступных гриду, с помощью семантических описаний компонент и данных, как это делается в Семантическом гриде, и службы обнаружения знаний.

Биомедицинские гриды должны вырабатывать, использовать и размещать знания как основу для современных приложений и будут базироваться, главным образом, на *гридах знаний* и *семантических гридах*. Используя службы высокого уровня, они сделают возможным предоставление пользователю информации, знаний, медицинских руководств и результатов исследований в доступной форме.

В этом направлении движется грид в области биомедицинской информатики, разработанный Национальным раковым институтом (National Cancer Institute – www.nci.nih.gov) и предназначенный для исследования раковых заболеваний (*caBIG* - Cancer Biomedical Informatics Grid). Инфраструктура *caBIG* свяжет источники данных, относящихся к раковым заболеваниям, инструментальные средства, отдельных людей и организации и поможет пересмотреть методику проведения исследований, оказание медицинской помощи и взаимодействие пациентов и партнеров по гриду с предприятиями, проводящими биомедицинские исследования (cabig.nci.nih.gov/caBIG/overview/).

Биомедицинские гриды могут оказать помощь в хранении, интеграции и анализе данных, получаемых или используемых в экспериментах и исследованиях (например, данных из общедоступных баз данных). Кроме того, они будут поддерживать моделирование, разработку и выполнение экспериментов с потоками работ (например, эксперименты "in silico"), используя стандартные способы моделирования, такие, как UML (Unified Modeling Language), онтологии и языки потоков работ. Главные концептуальные слои биомедицинских гридов включают в себя:

- *Слой источников данных и моделирования.* Данные, полученные в ходе экспериментов (например, масс-спектрометрия, микрочипы и т.д.), данные из общедоступных баз данных (например, PDB - Protein Data Bank, SwissProt) и данные клинической практики должны быть смоделированы с использованием уже устоявшихся и новейших методов управления знаниями, такими как UML и онтологии. Источники данных должны быть интегрированы, чтобы обеспечить легкий доступ к требуемой информации или к семантически связанным данным. Главные задачи этого слоя: базирующееся на онтологиях моделирование биологических / биомедицинских баз данных; моделирование распределенных биомедицинских приложений, таких как эксперименты “in silico”. Моделирование должно выполняться на всех этапах эксперимента, таких как подготовка образца, генерирование данных, предварительная обработка и фильтрование данных, анализ медицинских изображений, анализ данных биоинформатики, анализ биомедицинских данных, визуализация результатов [23].
- *Слой композиции и выполнения приложений.* Этот слой композиции потока работ предоставляет возможности для реализации сложных приложений биоинформатики и биомедицинских приложений (например, эксперименты *in silico*) посредством композиции открытых базовых инструментальных средств биоинформатики, которые будут выполняться в гриде с использованием ресурсов и данных, предоставленных исследовательскими центрами, входящими в различные виртуальные организации. На слое моделирования полезные программные средства должны быть классифицированы по отношению к технологии и аспектам их использования. Ключевыми вопросами данного слоя являются: онтологии предметной области для моделирования открытых компонент программного обеспечения биоинформатики и общедоступные биологические базы данных; базирующийся на онтологиях запрос и просмотр онтологий предметной области для поиска, отбора и локализации ресурсов биоинформатики и биомедицинских ресурсов (данные и компоненты программного обеспечения), которые должны использоваться для композиции приложений; моделирование и планирование выполнения распределенных приложений в гриде на базе потока работ; широкое использование открытых компонент программного обеспечения и компонент, разработанных в исследовательских центрах.
- *Слой анализа данных и извлечения знаний.* На этом слое развитые средства анализа данных, скомпонованные с использованием технологий потока работ, делают возможным извлечение знаний, полезных для проведения экспериментов. Этот слой должен содержать набор плагинов для анализа данных, использующих различные методы и подходы, например, статистический анализ и извлечение данных, анализ продолжительности существования и анализ других временных данных, визуализация многомерных данных, классификация данных и т.д. (например, Knowledge Grid [22], PROTEUS [24]).

7.6. ДОРОГА К ИСПОЛЬЗОВАНИЮ ГРИДА В ГЕНОМНОЙ МЕДИЦИНЕ

Грид-технология еще находится на ранней стадии развития. Дорога к её использованию еще не ясна, но можно разработать общую дорожную карту для успешного применения грид-технологии в области геномной медицины. Вот некоторые из необходимых шагов:

1. разработка специальных семантических грид-служб для среды интеграции данных;
2. установка и тестирование первых прототипов промежуточного программного обеспечения грида для сектора здравоохранения (исследование и медицинское обслуживание);
3. разработка, установка и тестирование первых грид-приложений для геномной медицины;
4. воспитание и распространение грид-культуры посредством обучения врачей, научных сотрудников и других людей, занятых в геномной медицине.

7.7. ССЫЛКИ

- [1] S. Collins and V. A. McKusick (2001) "Implications of the Human Genome Project for medical science", JAMA, (285): pp. 540-4.
- [2] D. J. Weatherall (2003) "Genomics and global health: time for a reappraisal", Science, (302): pp. 597-599.
- [3] A. Tefferi, M. E. Bolander, S. M. Ausell, E. D. Wieben and T. C. Spelsberg (2002) "Primer on medical genomics. Microarray experiments and data analysis", Mayo Clinical Proceedings, 77(9): pp. 972-940.
- [4] J. R. Nevins, E. S. Huang, H. Dressman, J. Pittman, A. T. Huang and M. West (2003) "Towards integrated clinico-genomic models for personalized medicine: combining gene expression signatures and clinical factors in breast cancer outcomes prediction", Human Molecular Genetics, (12): pp. 153-157.
- [5] K. K. Jain (2002) "Role of proteomics in diagnosis of cancer", Technological Cancer Research Treatments, 1(4): pp. 281-286.
- [6] J. Licinio and M.-L. Wong (Eds.) (2002) Pharmacogenomics: the search for Individualized therapies.
- [7] J. S. Ross, G. P. Linette, J. Stec, E. Clark, M. Ayers and N. Leschly (2004) "Breast cancer biomarker and molecular medicine", Expert Revisions in Molecular Diagnosis, 4(2): pp. 169-188.
- [8] A. Bayat (2002) "Science, medicine, and the future: Bioinformatics", British Medical Journal (BMJ), 324.
- [9] A. S. Pereira, V. Maojo, F. Martin-Sanchez, A. Babic and S. Goes (2002) "The INFOGENMED project" In ICBME 2002, Singapore.
- [10] F. Martin-Sanchez, V. Maojo and G. Lopez-Campos (2002) "Integrating Genomics into Health Information Systems", Methods of Information in Medicine, 41: pp. 25-30.
- [11] R. Sokolowski (1999) "Expressing Health Care Objects in XML" In EE 8th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Palo Alto, California, pp 341-342.
- [12] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000) "Gene Ontology: tool for the unification of biology", Nature Genetics, 25(25-29).

- [13] G. Aloisio, M. Cafaro, S. Fiore and M. Mirto (2004) "The GRelC Project: Towards Grid-DBMS" In Parallel and Distributed Computing and Networks (PDCN) - IASTED, Innsbruck, Austria.
- [14] BIOINFOMED (2003) Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare, BIOINFOMED Study, Report White Paper.
- [15] I. Foster, C. Kesselman and S. Tuecke (2001) "The Anatomy of the Grid: Enabling scalable virtual organizations", International Journal of High Performance Computing Applications, 15(3): pp. 200-222.
- [16] I. C. Oliveira, J. L. Oliveira, F. Martin-Sanchez, V. Maojo and A. S. Pereira (2004) "Biomedical information integration for health applications with Grid: a requirements perspective" In HealthGrid 2004, Clermont-Ferrand, France.
- [17] G. Berti, S. Benkner, J. W. Fenner, J. Fingberg, Lonsdale, S. E. Middleton and M. SurrIDGE (2003) "Medical Simulation Services via the Grid" In Nörager, S., Healy, J.-C. and Paindaveine, Y. (Eds) 1st European HealthGrid Conference, Lyon, France.
- [18] M. Cannataro and D. Talia (2004) "Semantic and Knowledge Grids: Building the Next-Generation Grid", IEEE Intelligent Systems (ISSI-0095-1203) - Special Issue on E-Science, 19(1): pp. 56-63.
- [19] N. Zhong and J. Liu (Eds.) (2004) Intelligent Technologies for Information Analysis, Springer Verlag (to appear).
- [20] T. R. Gruber (1993) "A translation approach to portable ontologies", Knowledge Acquisition, 5(2): pp. 199-220.
- [21] D. d. Roure, N. R. Jennings and N. Shadbolt (2003) "The Semantic Grid: A future e-Science infrastructure" in Grid Computing: Making The Global Infrastructure a Reality, (Eds.) Berman, F., Hey, A. J. G. and Fox, G., John Wiley & Sons:pp. 437-470.
- [22] M. Cannataro and D. Talia (2003) "KNOWLEDGE Grid An Architecture for Distributed Knowledge Discovery", CACM, 46(1): pp. 89-93.
- [23] C. F. Taylor (2003) "A systematic approach to modelling capturing and disseminating proteomics experimental data", Nature Biotechnology, 21: pp. 247-254.
- [24] M. Cannataro, C. Comito, F. Lo Schiavo and P. Veltri (2004) "Proteus, a Grid based Problem Solving Environment for Bioinformatics: Architecture and Experiments", IEEE Computational Intelligence Bulletin, 3(1): pp. 7-18.