The Gene Regulation Ontology (GRO):
- Design Principles and Use Cases -

Elena BEISSWANGER[a], Vivian LEE[b], Jung-Jae KIM[b], Dietrich REBHOLZ-SCHUHMANN[b], Andrea SPLENDIANI[c], Olivier DAMERON[c], Stefan SCHULZ[d], Udo HAHN[a]

[a]*Jena University Language and Information Engineering (JULIE) Lab, Jena, Germany*
[b]*European Bioinformatics Institute, Hinxton, Cambridge, UK*
[c]*Laboratoire d'Informatique Médicale, Université de Rennes 1, Rennes, France*
[d]*Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany*

05/26/2008

Gene Regulation and Regulatory Processes

- Gene expression
  - Synthesis of gene products (RNA and proteins)
  - Two steps: transcription and translation
  - Transcription: Gene → RNA
    (mediated by transcription factor proteins (TF) that regulate (up / down) the synthesis of RNA by a polymerase enzyme)
  - Translation: RNA → protein

- Regulation of Gene Expression
  - Control of the amount of gene products synthesized (at a particular time and under particular extra- and intracellular conditions)
  - Occurs during all steps of gene expression
  - Enables the cell to adapt to different conditions controlling its structure and function
  - Abnormal regulation may cause serious diseases

Rationale for a Gene Regulation Ontology

- Well-defined vocabulary for semantic annotations in scientific documents on gene regulation (EU BOOTStrep project)

- Semantically annotated text corpora as prerequisite for supervised machine learning algorithms

- Purpose: automatic population of a knowledge repository on gene regulation

Selected List of Gene Regulation Related Ontology Resources

Resource with URL	Relevant Information
Gene Ontology (GO) `http://geneontology.org/`	molecular functions, biological processes, cellular components
Sequence Ontology (SO) `http://sequenceontology.org/`	sequence regions and attributes of sequence regions
ChEBI `http://www.ebi.ac.uk/chebi/`	molecules of biological interest
INOH Molecule Role (IMR) `http://www.inoh.org/`	transcription factors and their functional domains
NCBI taxonomy, `http://130.14.29.110/Taxonomy/`	eukaryotes, prokaryotes
TransFac `http://www.gene-regulation.com/`	transcription factors, domains of transcription factors

What's Missing ...

- Principled and expressive representation of gene regulation proper
  - regulatory processes and participants involved (genes, transcripts, proteins)
  - Relationships between processes and participants
  - Formal, computable definitions
  - Common standardized description language (e.g. OWL)

Construction of the GRO

- Manual construction of the foundational structure
  - Integrating basic knowledge from text books and the UMLS
- Extension based on existing OBO ontologies
  - Screening of OBO ontologies (GO, SO, ChEBI, IMR, NCBI taxonomy) for entries related to gene regulation
  - Extraction and integration of these entries in GRO while keeping the references to the sources
- Extension based on domain specific databases
  - Integration of transcription factors entries extracted from the transcription factor database TransFac
- Extension based on literature screening
  - Analysis of 150 Medline abstracts (selected by a MeSH query and additional criteria) with regard to potentially new GRO terms

Size and Structure of the GRO

- Size (gro-v0.3)
  - 433 classes, 457 taxonomic relations
  - 8 relation types (+ inverses)
  - 404 class restrictions

- Bi-partite upper ontology
  - Continuant branch: entities 'which persist through time'
    - Physical continuant branch: entities having spatial dimension (e.g. gene, regulatory sequence, and protein)
    - Non-physical continuant branch: entities having no spatial dimension (e.g. protein function)
  - Occurrent branch: entities 'which have temporal parts'
    - e.g. transcription, gene expression, and various regulatory processes

- Represented in OWL DL

Relations in the GRO

- GRO classes highly interlinked by semantic relations
  - *partOf | hasPart* relating spatial or temporal parts to the whole
    - protein domain *partOf* protein, transcription initiation *partOf* transcription
  - *fromSpecies* relating species information
    - bacterial RNA polymerase *fromSpecies* bacterium
  - *participatesIn | hasParticipant* relating processes and events to the entities involved
    - and sub-relations *agentOf | hasAgent*, *patientOf | hasPatient*
    - regulation of transcription *hasAgent* transcription regulator
  - *encodes | encodedIn* relating genes to proteins
  - *functionOf | hasFunction* linking functions to their bearers
  - *hasQuality* specifies qualities inherent in particular entities
  - *resultsIn | resultsFrom* identifies the outcome of a process
  - *located-in | location-of*

Structure of GRO Classes
Example: Class TranscriptionFactor

CLASS EDITOR

For Class: TranscriptionFactor   (instance of owl:Class)   ☐ Inferred View

Property	Value
rdfs:comment	
definition	A transcription factor that binds to a specific DNA sequence in order to modulate transcription. The transcription factor may or may not also interact selectively with a protein (other transcription factors of cofactors) or protein or macromolecular complex. (modified from [GO:0003700])
rdfs:label	transcription factor
synonym	gene regulatory protein

Asserted

NECESSARY

TranscriptionRegulator

hasFunction **some** TranscriptionFactorActivity

OWL class restriction

hasPart **some** AminoAcid   [from Pr

Vocabulary for Semantic Annotation of Scientific Documents

- Semantic annotations on two levels:
  1. Annotation of terms denoting continuants (e.g., transcription factor proteins and genes)
     - Vocabulary: terms from the GRO continuant branch
  2. Annotation of regulatory processes / event annotation
     - Much more complex task, requires annotation of continuants (1.)
     - Vocabulary: terms from the GRO occurrent branch
     - Participation relations specified for these terms are exploited to constrain semantic roles.

Ontology Classes as a Vocabulary for Semantic Annotation

Characterization of the regulon controlled by the leucine-responsive regulatory protein in Escherichia coli.

The leucine-responsive regulatory protein (Lrp) has been shown to regulate, either positively or negatively, the transcription of several Escherichia coli genes in response to leucine. We have used two-dimensional gel electrophoresis to analyze the patterns of polypeptide expression in isogenic lrp+ and lrp mutant strains in the presence or absence of leucine. The absence of a functional Lrp protein alters the expression of at least 30 polypeptides. The expression of the majority of these polypeptides is not affected by the presence or absence of 10 mM exogenous leucine.

transcription factor
ligand (chemical entity)
nucleotide sequence
experimental intervention
regulatory process
transcription
gene expression

Ontology Classes as a Vocabulary for Semantic Annotation

Characterization of the regulon controlled by the leucine-responsive regulatory protein in Escherichia coli.

The leucine-responsive regulatory protein (Lrp) has been shown to regulate, either positively or negatively, the transcription of several Escherichia coli genes in response to leucine. We have used two-dimensional gel electrophoresis to analyze the patterns of polypeptide expression in isogenic lrp+ and lrp mutant strains in the presence or absence of leucine. The absence of a functional Lrp protein alters the expression of at least 30 polypeptides. The expression of the majority of these polypeptides is not affected by the presence or absence of 10 mM exogenous leucine.

transcription factor
ligand (chemical entity)
nucleotide sequence
experimental intervention
regulatory process
transcription
gene expression

SWRL Rules

- Rules in Semantic Web Rule Language (SWRL) were defined on GRO classes and relations

- Help to refine event classification in text

- Example:
  - Given that a reference to a GeneRegulation event has been identified in text during the annotation step
  - … and given appropriate other events and participants have been identified
  - …   a SWRL rule that has been defined based on GRO allows to infer that a GeneRegulation event is in fact a TranscriptionRegulation event (more specific)

SWRL Rules – an Example
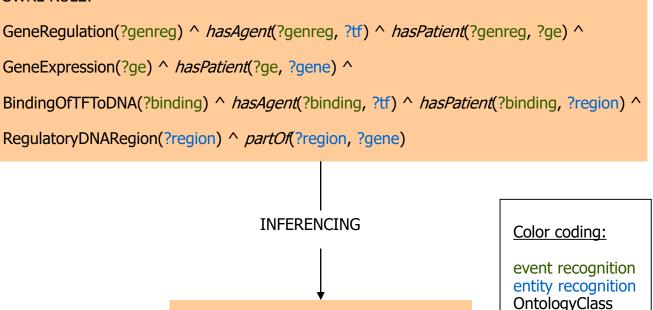
FROM GRO:

GeneRegulation
  hasAgent TranscriptionFactor
  hasPatient GeneExpression

GeneExpression
  hasPatient Gene

TranscriptionRegulation
  isA GeneRegulation
  hasAgent TranscriptionFactor
  hasPatient Gene

BindingOfTFToDNA
  hasAgent TranscriptionFactor
  hasPatient RegulatoryRegion

RegulatoryRegion
  partOf Gene

SWRL RULE:

GeneRegulation(?genreg) ^ *hasAgent*(?genreg, ?tf) ^ *hasPatient*(?genreg, ?ge) ^

GeneExpression(?ge) ^ *hasPatient*(?ge, ?gene) ^

BindingOfTFToDNA(?binding) ^ *hasAgent*(?binding, ?tf) ^ *hasPatient*(?binding, ?region) ^

RegulatoryDNARegion(?region) ^ *partOf*(?region, ?gene)

INFERENCING

TranscriptionRegulation(?genreg)

Color coding:

event recognition
entity recognition
OntologyClass
ontologyRelation

Availability of GRO

- GRO is freely availabe
- GRO website:
  - **http://www.ebi.ac.uk/Rebholz-srv/GRO/GRO.html**
- Access to GRO via the OBO library:
  - **http://www.obofoundry.org/**
  - (see section 'Other ontologies and terminologies of interest')
- Access to GRO via the NCBO BioPortal:
  - **http://www.bioontology.org/ncbo/faces/pages/ontology_list.xhtml**

Acknowledgements