Last updated: 22-06-01

Axiomatization and refactoring of the ‘chromosomal disorder’ branch of Mondo.

Sabrina Toro[1], Nicole Vasilevsky[1], Nico Matentzoglu[2], Peter N Robinson[3], Christopher Mungall[4]

1. University of Colorado Denver, Anschutz Medical Campus, Aurora, CO USA
2. Semanticly, Athens, Greece
3. Jackson Laboratory, Farmington, CT, USA
4. Lawrence Berkeley National Laboratory, Berkeley, CA, USA

See GitHub issues:
https://github.com/monarch-initiative/mondo/issues?q=is%3Aissue+label%3A%22Chr+anomaly%2Fmonochrom%22+

1. General considerations

- **Monochrom :**
  - Monochrom was created to convert chromosomes and chromosome bands data from UCSC Genome Browser into an OWL classification, allowing for these terms to be used in axioms.
  - See more here: https://github.com/monarch-initiative/monochrom

- **Goal of the project:**
  - Axiomizing ‘chromosomal disorder’ using Monochrom terms in the logical definition provide a relationship to the chromosome or chromosome part affected
  - Review the ‘chromosomal disorder’ branch
    - Classification was inconsistent (e.g. ‘autosomal anomaly’ and ‘gonosomal anomaly’ were organized differently)
    - Concepts were mixed up (e.g. terms referring to actual chromosomal change versus terms referring to disease)

- **Axioms**:
  - The type of chromosomal change (e.g deletion of a part of a chromosome) is expressed with the axiom :
    `‘has modifier’ some chromosome_variation (SO:0000240)`
  - The chromosome or chromosomal part that is affected (and is the root of the disease) is expressed with the axiom:
    `‘disease arises from structure’ some (chromosome (GO:0005694) or ‘chromosomal region’ (GO:0098687))`

The terms referring to ‘chromosome’ and ‘chromosome_part’ come from Monochrom.

- **Concepts**:
  Reorganization of the ‘chromosomal disorder’ branch was based on these high level concepts:
  - Chromosome number anomaly
    - Anomaly due to a change in the number of chromosomes. This is referring to a change in the ENTIRE chromosome (not partial chromosome)
    - Includes aneuploidy (monosomy, trisomy, tetrasomy,...) and polysomy.
  - Chromosome structure anomaly
    - Includes changes in the structure of a chromosome, such as partial duplication/deletion, ring chromosome, as well as complex structural anomaly (e.g. inversion and insertion,...)
  - Anomaly based on Chromosome type
    - It is useful to have anomalies grouped by the Chromosome affected (e.g. Down syndrome)
  - Other chromosomal anomalies, such as disomy do not fit any of the categories above

  Other concepts:
  - **Mosaic versus complete**: refers to an anomaly which is found in some cells of an individual, or in all the cells of an individual, respectively.
    - See section 5
  - **‘Partial’ versus ‘total’**: most often refers to deletion/duplication of chromosomes, and whether part of the chromosome or the full chromosome, respectively. The terms “partial” and “total” are often used in the community. In Mondo, these are referred to ‘partial deletion/duplication of chromosome’ and ‘aneuploidy’, respectively.

- ‘chromosomal disorder’
  - We did <u>not</u> add the following axiom to the parent term “chromosomal disorder”, as we did not think we needed to add this specificity: `'disease arises from structure' some (chromosome or chromosome_part)`

Last updated: 22-06-01

2. Main concept: Chromosome number

2.1. Concept: aneuploidy

- Definition: A chromosomal abnormality in which there is an addition or loss of chromosomes within a set. [NCIT:C2873]
    - This term is different from the loss or addition of a part of the chromosome
- Equivalence axiom:
  `'chromosomal disorder' and ('has modifier' some aneuploid)`
  in which 'aneuploid' refers to SO:0000054

2.1.1. Concept: monosomy

- Definition: A chromosomal abnormality consisting of the absence of one chromosome from the normal diploid number. [NCIT:C3239]
    - This term refers to the absence of an entire chromosome (not part of a chromosome
    - = 'total' monosomy
- Equivalent:
  `aneuploidy and ('has modifier' some hypoploid)`
  in which 'hypoploid' refers to SO:0000056
- Children of this term are defined by specifying which chromosome is absent using a Monochrom 'chromosome' term. For example: 'monosomy X' (MONDO:0020466) :
  `monosomy and ('disease arises from structure' some 'chrX (Human)')`
- Pattern:
  https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/monosomy

2.1.2. Concept: trisomy

- Definition: A chromosomal abnormality consisting of the presence of one chromosome in addition to the normal diploid number. [NCIT:C3421]
    - This term refers to the addition of an entire chromosome (not part of a chromosome)
    - = 'total' trisomy
- Equivalent: this term cannot be defined using an "equivalent" axiom because there is currently no SO term to distinguish a trisomy from a 'hyperploidy' (aneuploidy which has extra number of chromosomes of a same type). However, this term and its children can have the SubclassOf axiom: `('has modifier' some hyperploid)`
  `hyperploid = SO:0000055`

- Children of this term are defined by specifying which chromosome is duplicated using a Monochrom 'chromosome' term. For example: 'chromosome 16 trisomy' (MONDO:0022180) : `trisomy and ('disease arises from structure' some 'chr16 (Human)')`
- Pattern: https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/trisomy

2.1.3. Concept: tetrasomy

- Definition: A chromosomal disorder consisting of the presence of two chromosomes of the same type in addition to the normal diploid number.
    - This term refers to the addition of 2 entire chromosomes of the same type (not part of a chromosome)
    - = 'total' tetrasomy
- Equivalent: this term cannot be defined using an "equivalent" axiom because there is currently no SO term to distinguish a tetrasomy from a 'hyperploidy' (aneuploidy which has extra number of chromosomes of a same type). However, this term and its children can have the SubclassOf axiom: `('has modifier' some hyperploid)` in which 'hyperploid' refers to SO:0000055
- Children of this terms are defined by specifying which chromosome is in 2 extra copies using a Monochrom 'chromosome' term. For example: 'tetrasomy 21' (MONDO:0019864) : `tetrasomy and ('disease arises from structure' some 'chr21 (Human)')`
- Pattern: https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/tetrasomy

2.1.4. Concept: pentasomy

- Definition: A chromosomal disorder consisting of the presence of three chromosomes of the same type in addition to the normal diploid number.
    - This term refers to the addition of 3 entire chromosomes of the same type (not part of a chromosome)
- Equivalent: this term cannot be defined using an "equivalent" axiom because there is currently no SO term to distinguish a tetrasomy from a 'hyperploidy' (aneuploidy which has extra number of chromosomes of a same type). However, this term and its children can have the SubclassOf axiom: `('has modifier' some hyperploid)` in which 'hyperploid' refers to SO:0000055
- Children of this term are defined by specifying which chromosome is in 3 extra copies, using a Monochrom 'chromosome' term. For example: 'pentasomy X'

(MONDO:0015228) : `pentasomy and ('disease arises from structure' some 'chrX (Human)')`

- Pattern: https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/pentasomy

2.2. Concept: polyploidy

- Definition: The chromosomal constitution of a cell containing multiples of the normal number of chromosomes; includes triploidy (symbol: 3N), tetraploidy (symbol: 4N), etc. [MESH:D011123]
- Polyploidy refers to a chromosomal disorder in which cells contain multiples of the normal number of chromosomes; includes triploidy (symbol: 3N), tetraploidy (symbol: 4N), etc.
- Equivalent: `'chromosomal disorder' and ('has modifier' some polyploid)` in which 'polyploidy' refers to SO:0001254

3. Main concept: Chromosome structure

3.1. Concept: ring chromosome

- A ring chromosome is an aberrant chromosome whose ends have fused together to form a ring.
- Parent term (newly created) **'ring chromosome disorder** (MONDO:0700091) was defined with the Equivalent axiom:
  `'chromosomal disorder' and ('has modifier' some ring_chromosome)`
  in which 'ring_chromosome' refers to SO:1000045
  Note: The term 'ring chromosome' (MONDO:0018186) was an orphanet grouping term and child of "autosomal monosomy". We created a new term to refer to a more general concept of 'ring chromosome' which includes autosomes and gonosomes; MONDO:0018186 was obsoleted.
- Children of 'ring chromosome disorder' were defined with the Equivalent axiom:
  `'ring chromosome anomaly' and ('disease arises from structure' some chromosome)` in which 'chromosome' is a Monochrom term that refers to the chromosome affected.
- Pattern: https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/ring_chromosome_anomaly

3.2. Concept: partial chromosome deletion

- Refers to the deletion of part of a chromosome. Some groups call these “partial monosomy”, or “monosomy 10q” (where the name of the chromosome is a chromosomal region.
- Axioms:
  - `‘has modifier’ some chromosomal_deletion (SO:1000029)`
  - `‘disease arises from structure’ some (‘chromosomal region’ (= GO:0098687)or chromosome (=GO:0005694))`
- Parent term: ‘syndrome caused by partial chromosomal deletion’ (MONDO:0000761)
  - Equivalent: `‘chromosomal disorder’ and (‘has modifier’ some chromosomal_deletion)`
- Children of ‘syndrome caused by partial chromosomal deletion’ were defined with the Equivalent axiom:
  `'syndrome caused by partial chromosomal deletion' and (‘disease arises from structure’ some (‘chromosomal region’ or chromosome)` in which ‘chromosome’ and ‘chromosomal region are a Monochrom term that refers to the chromosome/chromosomal region affected.
- Pattern for children terms:
  - chromosomal region deleted is known: https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/chromosomal_region_deletion.yaml
  - chromosomal region is not known, but the chromosome is known https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/partial_chromosomal_deletion.yaml
- Note:
  - Some terms refer to the “distal” or “non-distal” region of the chromosomal region being deleted. For example: ‘distal 10q deletion syndrome’ (MONDO:0012315), ‘non-distal monosomy 10q’ (MONDO:0015546)
    In these cases, no equivalence axiom was added, and the following subclassOf axioms were added:
    - `‘disease arises from structure’ some ‘chromosomal region’`
    - `‘has modifier” some ‘chromosomal_deletion’`

3.3. Concept: partial chromosome duplication

- Refers to the duplication of part of a chromosome. Some groups call these “partial trisomy”, or “trisomy 10q” (where the name of the chromosome is a chromosomal region).
- Axioms:
  - `‘disease arises from structure’ some ‘chromosomal region’ (= GO:0098687)`
- Parent term: ‘syndrome caused by partial chromosomal duplication’ (MONDO:0000762)

    - Definition: A chromosomal disorder consisting of the presence of a part of a chromosome in more copies than in a regular genome.
        - This definition includes this term includes partial trisomy (one additional copy), and partial tetrasomy (2 additional copies) of a chromosome section.
    - ‘chromosomal_duplication’ (SO: 1000037)
      Definition: An extra chromosome.
        This definition implies that the duplication involves a full extra chromosome, however, the classification of this term in SO inferred that this term actually refers to a partial chromosomal duplication (e.g. this term includes inter- and intra-chromosomal duplication.)
      Because of this inconsistency, we decided to refrain from using a SO term as modifier until SO updated these terms
- Children terms Equivalent:
  `‘syndrome caused by partial chromosomal duplication’ and (‘disease arises from structure’ some (‘chromosomal region’ or chromosome)`
  in which ‘chromosome’ and ‘chromosomal region are a Monochrom term that refers to the chromosome/chromosomal region affected.
- Pattern for children terms:
    - http://purl.obolibrary.org/obo/mondo/patterns/chromosomal_region_duplication.yaml

4. Main concept: uniparental disomy

- These terms were classified under ‘autosomal uniparental disomy’ (MONDO:0020055) (under ‘autosomal anomaly’, MONDO:0020049), ‘uniparental disomy of chromosome X’ (MONDO:0017011) (under ‘gonosome anomaly’, MONDO:0020058)
- A new grouping term (chromosome agnostic) was created: **‘uniparental disomy’** (MONDO:0700086), defined as “a condition characterized by the inheritance of a chromosome pair from one parent and no chromosomal copies from the other parent. It results in developmental abnormalities or rare recessive disorders.” [NCIT:C85215]
- The existing ‘uniparental disomy’ terms were made children of this new term, with the affected chromosome being reported with the axiom
  `‘disease arises from structure’ some chromosome`
  in which ‘chromosome’ is a Monochrom term that refers to the chromosome affected.
- Note that we did not create grouping terms to refer to maternal or paternal origin (‘uniparental disomy of maternal origin’, ‘uniparental disomy of paternal origin’)

Last updated: 22-06-01

5. Main concept: mosaic vs complete

- We created new disease characteristics:
  - ‘mosaic’ (MONDO:0700062) refers to a disease characteristic in which the cause of the disease is present in some of the cells of the organism.
  - ‘complete’ (MONDO:0700063) refers to a disease characteristic in which the cause of the disease is present in all the cells of the organism.
- These characteristics can be used in axioms with the relation:
  - `‘has modifier’ some mosaic`
  - `‘has modifier’ some complete`
- Examples: complete trisomy 13 (MONDO:0700033), mosaic trisomy 13 (MONDO:0700034)

6. Main concept : chromosome type autosomal/gonosomal

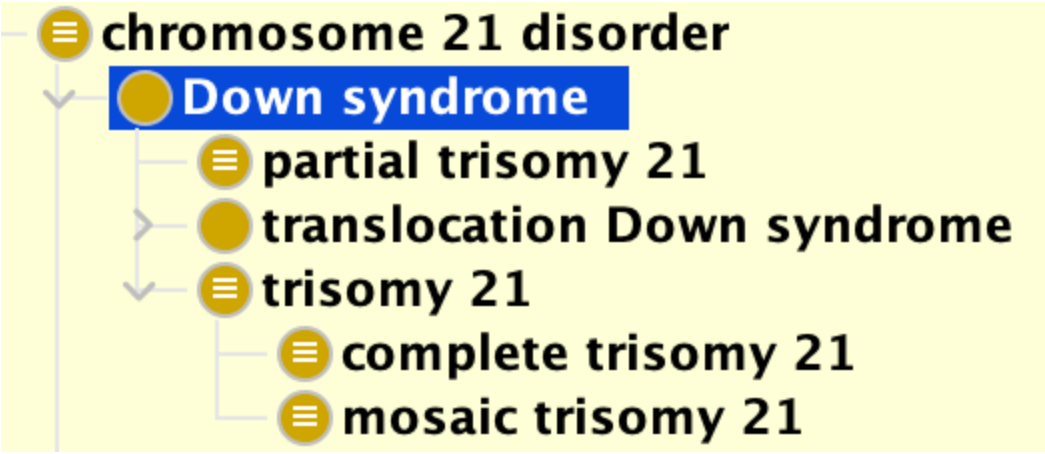
- Refers to chromosomal disorder based on the chromosome affected.
- Could be “autosomal” or “gonosomal” depending on whether the affected chromosome is a not sexual or sexual (X, Y) chromosome, respectively
- Axioms:
  `‘disease arises from structure’ some chromosome`
  in which ‘chromosome’ is a Monochrom term that refers to the chromosome affected.
- Pattern for children terms:
  https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/chromosome_type.yaml

7. Example:

7.1. Down syndrome

The term ‘Down Syndrome’ refers to the triplication of the whole or part of Chr21 with a “mosaic” and “complete” version of the disease.
The Mondo classification distinguishes between triplication of a *whole* chromosome (‘trisomy’ (MONDO:0700065) and the triplication *part o*f a chromosome (‘syndrome caused by partial chromosomal deletion’ (MONDO:0000761). Therefore, “Down syndrome” is a grouping term for ‘trisomy 21’ (MONDO:0700126) and ‘partial trisomy 21 (MONDO:0700130).
It should be noted that this classification is different from classification in other resources.

Last updated: 22-06-01



8. Challenges:

8.1. Orphanet classification mixed with other classification

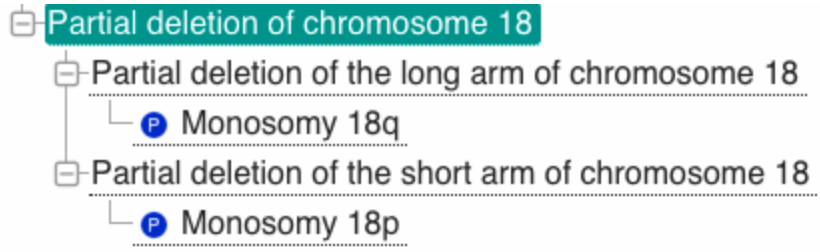
Disclaimer: At the time of this project, Mondo integrated Orphanet v3.3. Since this project came to completion, a new version of Orphanet (4.0) was released and is in the process of being synchronized with Mondo. Therefore some of the information in this section might be out of date.

Several Mondo terms are mapped to Orphanet grouping classes, and have only an Orphanet ID and no other database cross reference (x-ref). It should be noted that Orphanet had a different classification approach compared to Mondo whereby they often grouped classes by phenotypic features or very specific disease features.

- Issue #1: the Mondo terms mapped to these Orphanet grouping terms cannot be reused as general grouping terms since these Orphanet terms are too specific. We obsoleted the grouping terms which did not conform to our classification strategy.
    - For example: the term ‘ring chromosome’ (MONDO:0018186) was an Orphanet grouping term and child of “autosomal monosomy”. Therefore, in order to group all the ‘ring chromosome disorder’ terms (including terms refering to non-autosome), we had to create a new, more general term.
    - See list of obsoleted grouping classes here: https://github.com/monarch-initiative/mondo/issues/4272

- Issue #2 : Some Orphanet terms represent grouping terms that do not represent specific diseases. For example:

Last updated: 22-06-01



In this example, “Monosomy 18q” and “Monosomy 18p” represent actual diseases, and have a *part-of* relation to ‘Partial deletion of the long/short arm of chromosome 18’ which represent a grouping term and not an actual disease. We reviewed these terms, and on a case-by-case basis (taking into consideration other x-refs), we either obsoleted or merged these grouping terms.
See a list of some of the merged terms here:
https://github.com/monarch-initiative/mondo/issues/4278

Note: since completion of this chromosomal disorder reclassification project, an effort is currently underway to sync Mondo with the current version of Orphanet (v4.0) and many of these classifications have been since revised in Orphanet (see https://github.com/monarch-initiative/mondo/issues/4579).

8.2. Some terms do not refer to diseases (e.g. NCIT, MESH term)

Some terms, such as ‘ring chromosome’ (MONDO:0018186) and chromosome inversion (MONDO:0043678) refer to **a type of chromosome rearrangement**, describing the chromosomal change rather than an actual disease/syndrome. Many term labels and definitions were reviewed and renamed to ensure that they clearly represent a disease concept. In many cases, however, NCIT and MESH ids were retained as they are useful to users.

8.3. Chromosome region range

Some chromosomal disorders arise from a change in a **range of chromosome regions (**for example, ‘chromosome 1q41-q42 deletion syndrome’ (MONDO:0012927)) which did not exist in Monochrom. Since the number of terms requiring chromosome region range was limited (less than 40), we created new Monochrom terms for the chromosome region ranges.

9. Conclusion and future work:

The ‘chromosomal disorder’ branch of Mondo was reviewed and reclassified based on the type of chromosomal anomaly and the chromosome or chromosomal region affected in the disorder. Axiomatization of these terms using Sequence Ontology (SO) and Monochrom allows for better autoclassification and interoperability of terms between resources.

Last updated: 22-06-01

Terms from the ‘chromosome_variation’ (SO:0000240) branch of the SO were used to refer to the type of chromosomal anomaly at the root of the disorder. Future work will address revisions to the SO branch, as many terms are missing or ambiguously or erroneously defined. We reported issues to the SO tracker, and are planning to add axioms including reference to the ‘chromosome_variation’ terms currently missing in Mondo once these issues are resolved.

Monochrom was created to convert chromosomes and chromosome bands data from UCSC Genome Browser into an OWL classification.

Here, we reported our approach to axiomatize ‘chromosomal disorder’, in which the chromosomal anomaly is inherited. A similar approach could be taken for terms referring to chromosomal anomaly occurring in somatic cells, as is often the case in cancers, for example ‘acute myeloid leukemia, del(5q31-q32)’ (MONDO:0100401), ‘acute myeloid leukemia, Trisomy 8’ (MONDO:0100389). Axiomatizing diseases due to chromosal anomaly in somatic cells is an ongoing project.

This work was presented at the 14th annual biocuration conference (https://www.biocuration.org/14th-annual-biocuration-conference-virtual): abstract can be found https://zenodo.org/record/5570276#.Yj5fHBDMJcg (page 21), and poster https://doi.org/10.7490/f1000research.1118805.1.