HL7 v3 Clinical Genomics –
Overview

The HL7 Clinical Genomics Work Group

Prepared by Amnon Shabo (Shvo), PhD

HL7 Clinical Genomics WG
Co-chair and Modeling Facilitator

HL7 Structured Documents WG
CDA Co-editor
CCD Implementation Guide Co-editor
GTR Implementation Guide prime editor

HL7 RIMBAA WG, Co-chair
The Mission of HL7 Clinical Genomics Work Group

- The HL7 Clinical Genomics Work Group (CGWG) supports the HL7 mission to create and promote its standards by enabling the communication between interested parties of clinical and genomic data related to an individual. The focus of the CGWG efforts is the personalization of the genomic data, the so-called ’omics differences in an individual’s genomic – and its association with relevant phenotypic and clinical information. Associations to interpretive/expected phenotypes will be modeled as knowledge that can be utilized to transform an individual's data into meaningful information.

- CGWG will facilitate the development of common standards for clinical research information management across a variety of organizations -- including national and international government agencies and regulatory bodies, private research efforts, and sponsored research -- and thus the availability of safe and effective therapies by improving the processes and efficiencies associated with regulated clinical research.

- CGWG will strive to achieve common semantics across the clinical and research environments. Consequently, the group will start each standardization effort in Universal specifications that later on can be refined to specific realms.
Overview of Activities

Three Tracks:

v3:
- Family History (Pedigree) Topic
- Genetic Variations Topic
- Gene Expression Topic
- CMETs defined by the Domain

v2:
- v2 Implementation Guides
  - The IG “Genetic Test Result Reporting to EHR” is modeled after the HL7 Version 2.5.1 Implementation Guide: Orders And Observations; Interoperable Laboratory Result Reporting To EHR (US Realm), Release 1

CDA:
- A CDA Implementation Guide for Genetic Testing Reports

Common:
- Domain Analysis Models for the various topics (e.g., gene.exp, c.seq)
- A Domain Information Model (v3) describing the common semantics
- Semantic alignment among the various specs
HL7 Clinical Genomics v3 Static Models

- Family History
- Genetic Locus (DSTU Expired)
- Genetic Variation
- Gene Expression
- Phenotype (utilizing the HL7 Clinical Statement)
To achieve semantic interoperability...

...we need standard specs derived from a Central Health RIM:

Central Health RIM (e.g., an extended HL7 V3 Reference Information Model):
Bio & medical-informatics standard specs are derived from the same RIM
The DSTU Genetic Locus Model (deprecated) Focal Areas:

- Expression Data
- The Locus and its Alleles
- Sequence and Proteomics
- Sequence Variations
- (Clinical) Phenotypes
The Underlying Paradigm: Encapsulate & Bubble-up

Genomic Data Sources

Bridging is the challenge...

Clinical Practices

EHR System

Knowledge
(KBs, Ontologies, registries, reference DBs, Papers, etc.)

HL7 CG Messages with encapsulated data associated with HL7 clinical objects (phenotypes)

Decision Support Applications

Bubble up the most clinically-significant raw genomic data into specialized HL7 objects and link them with clinical data from the patient EHR

Encapsulation by predefined & constrained bioinformatics schemas

Bubbling-up is done continuously by specialized DS applications

Encapsulating HL7 Objects

HL7 CG Messages with mainly Encapsulating HL7 Objects
Encapsulate & Bubble-up Example

DNA Lab

Genetic Counseling

Sequencing Example...

HL7 CG Messages with a Sequence
Encapsulation by
a constrained
BSML schema

HL7 Object encapsulating the raw
sequencing results

Knowledge Sources
on genetic variants
(e.g., OMIM)

EHR System

Decision Support Applications
Bubble up the most clinically-significant SNP data into
HL7 SNP and Mutation objects and
link them with clinical data from the patient EHR

Bubbling-up is done dynamically
by specialized applications, e.g.,
sequence analyzing programs
Example: Family History XML Encoding

Taken from a patient pedigree, the portion related to patient’s daughter
(in collaboration with Partners HealthCare & other HL7 CG SIG members)

```
<!-- DAUGHTER -->
- <relationshipHolder>
  <id extension="555.011" />
  <code code="DAU" />
  + <relationshipHolder>
    +/- GENOMIC DATA -->
    - <subjectOf>
      - <clinicalGenomicChoice>
      - <clinicalGenomicChoiceGenotype>
        - <Genotype>
          - <individualAllele>
            <code code="BRCA1" codeSystem="[insert GenBank OID]"
              codeSystemName="GenBank" />
            <text>Homo sapiens breast and ovarian cancer susceptibility (BRCA1)
              complete cds.</text>
          </Genotype>
          </clinicalGenomicChoiceGenotype>
      </clinicalGenomicChoice>
  </subjectOf>
  +/- CLINICAL DATA -->
  + <subjectOf>
  </relationshipHolder>
  +/- end of DAUGHTER data -->
```

Point back to the raw data of this relative providing “personal evidence”
XML Fusion: Encapsulation of Raw Genomic Data

Raw genomic data represented in Bioinformatics markup

```
<subjectOf2>
    <geneticLocus>
        <component1>
            <individualAllele moodCode="EVN">
                <text>breast cancer 1, early onset</text>
                <value code="83990" displayName="BRCA1" codeSystemName="NCBI Entrez">
                    <translation code="20473" displayName="BRCA1" codeSystem="HGNC"/>
                </value>
            </individualAllele>
        </component1>
        <component2>
            <sequence moodCode="EVN">
                <code code="BSMLcon3"/>
                <value mediaType="text/xml">
                    <bsml:BSML xmlns:bsml="urn:bsml.org">
                        <bsml:Definitions>
                            <bsml:Sequences>
                                <bsml:Sequence id="seq1" molecule="dna" ic-acckey="U14680 REGION: 101..199" db-source="GenBank" title="BRCA1, exon 2" representation="raw" local-acckey="this could be used by the genetic lab">
                                    GCTCCCA CTCCATGAGG TATTTCTCA
                                    CATCCGTGC CCAGCCCGGC CGCGGGGAGC CCGCTTCAT CGCCGTGGGC
                                    TACGTCGAGC ACAACGAGTT CTGACGCTTC GACAGGAGAC CGCGGAGCCA
                                    GAGATTAGCA CCGGAGGCC GCTGCATAGA GAGGAGGAGG CGCGGAGTT
                                    GGGACACGGA GAACAGGAA ATGTAAGGGCC AGTCACAGAC TGACCGAGTG
                                    GACCTGGGGA CCCTGCAGCG CTACTACAAC CAGAGCAGGG CCG
                                </bsml:Sequence>
                            </bsml:Sequences>
                            <bsml:Sequence id="seq2" molecule="dna" ic-acckey="U14680 REGION: 200..253" db-source="GenBank" title="BRCA1, exon 3" representation="raw" local-acckey="this could be used by the genetic lab">
                                    GTTCTCA
                                    CACCATCCAG ATATATATAG GCTGCAGGCT GGCTGGGAGC GGGCGCTTCC
                                    TCCGGGGTA CCGCGAGGAC GCTGAGCGA GCGAGGATTAT CACGGCCCTT
                                    AACAGAGGCAC TGCTGCTTGA GACCGCCGGG GACATGGCG CTGACATCAC
                                    CAAGCCGAGT GGCGAGGGCC CCGATGTCGG GGAGCAGCAAG AGAGGCTACC
                                    TGGATTGGGAC GTGGTCGGAG TGCGTTCCAG GATACCTGGA GAACCGGGAAG
                                    GAGACGCTGAC AGCGCAAGG
                                </bsml:Sequence>
                            </bsml:Sequences>
                        </bsml:Definitions>
                    </bsml:BSML>
                </value>
            </sequence>
        </component2>
    </geneticLocus>
</subjectOf2>
```
CMET – a shared v3 model used by domain models

Current CMET efforts:
- Phenotype
- Genetic Variation
- Gene Expression
- Pedigree
The Phenotype CMET Model

HL7 Clinical Genomics SIG
Document: Genotype Topic - The Phenotype Model
Rev: COCT_RM340000UU (Phenotype-v13) Date: March 16, 2008
Facilitator: Amnon Shebo (Shvo), IBMResearch in Haifa, shebo@il.ibm.com

Phenotype (COCT_RM340000UU)
Entry point to the Phenotype CMET, used by the Clinical Genomics models to describe complex phenotypes associated with genomic observations. In case of interpretive phenotypes, the model is used to represent complex phenotypes that cannot be represented by a single code in the InterpretationCode attribute of the source genomic observation.

Note:
Phenotype (COCT_RM340000UU)
Entry point to the Phenotype CMET, used by the Clinical Genomics models to describe complex phenotypes associated with genomic observations. In case of interpretive phenotypes, the model is used to represent complex phenotypes that cannot be represented by a single code in the InterpretationCode attribute of the source genomic observation.

Observed Phenotype

Interpretive Phenotype

Note:
Use this CMET to describe a complex phenotype (e.g., adverse drug reaction in the context of a specific mutation) that is embedded in instances compliant with this model and not just referenced. The complex phenotype is either known in the scientific literature as a possible phenotype, or has been actually observed in the patient.

Note:
Use this CMET to describe medication components related to this phenotype (e.g., in pharmacogenomics use), when the administration of the drug is described elsewhere.
The Genetic Variation CMET (passed normative in Jan. 2010)

<table>
<thead>
<tr>
<th>Genetic Loci</th>
<th>Genomic Observations</th>
<th>Interpretive Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (including specimen)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associated data (vocab. Controlled)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed or Interpretive phenotypes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing issues: collecting specimen, extracting genetic material, identifying genomic observations, interpretation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genetic Report (CDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic Locus</td>
</tr>
<tr>
<td>Individual Allele</td>
</tr>
<tr>
<td>Sequence Variation</td>
</tr>
<tr>
<td>Sequence (observed or reference)</td>
</tr>
<tr>
<td>Genetic Testing Order</td>
</tr>
</tbody>
</table>
The Gene Expression CMET Draft

- Genetic Loci
- GTR Report
- Genetic Locus
- Gene Expression
- Participants
- Associated observations
Family History: PHR-EHR-GEN Convergence

Enable
Decision Support
e.g., risk analysis algorithms

PHR

EHR

Genomics

Population Research and Public Health

* Detailed description can be found in the Family History presentation
The Pedigree Spec - Updates

- HL7 Normative Standard since 2007 and reaffirmed in May 2012 for 5 more years

- Pilots:
  - Hughes Risk Applications
  - Surgeon General’s My Family Health Portrait
  - Many more…

- HL7 US Realm IG
  - Under development with expected balloting in Sept 2012
  - Transmission supported in CCD
  - Accommodate MU Stage 2 requirements

- Release 2 in progress
  - Update representation of genetic data
  - Accommodating ISO ballot comments

- FHIR Pedigree?!
Pedigree Implementation Guide for US Realm

- Specify the exchange format to be compliant with MU
  - Reusing HITSP CCD exchange; CCD instance references a Pedigree

- Simplify the genetic representation by constraining out the full blown genomic CMETs and use GTR suggested conventions

- Specify clinical data using SNOMED (following HughesRiskApp IG)
  - Do we need dynamic binding?

- Representing interpretation as in GTR (constrain out the interpretation code attribute)

- Pedigree R1 was adopted by MU 2 Final Ruling
  "The HL7 Pedigree standard was originally released in 2007. Release 1 was recently reaffirmed by the American National Standards Institute (ANSI), which is a process that occurs every five years. We have adopted this reaffirmed version as it is the same version (Release 1) of the standard as the version we proposed. An implementation guide for this standard is scheduled to be published shortly after this final rule. Although EHR technology will not be required to conform to the implementation guide for certification, the implementation guide will provide important guidance for use of the HL7 Pedigree standard with EHR technology."

- Reorganize the guide to have two parts:
  - Basic
    - The backbone of links among relatives & proband
    - Demographics
    - Clinical data
  - Advanced
    - Genetic data
    - Risk assessment results
FHIR Pedigree

- **Scoping:**
  - Risk assessment results
  - Genetic models
  - Branching out to any no. of generation

- Backward compatible?

- Mapped back to the v3 model!

- Do we have the bandwidth to do it?

- We’re asked on no. and scope of genomic resources
Other v3 Models / Efforts

- The Genome
  - An HL7 Domain Information Model (DIM)
  - Evolved as generalization of the Genotype DSTU models
  - Was not balloted
  - Will be updated based on the DAM efforts
  - In the future, DAM balloting will require the mapping to DIM

- The RCRIM use of our DSTU CMETs

- Query model
The Domain Information Model - Genome

Individual Allele

Expression Attributes

Expression Data

Sequence Variation (SNP, Mutation, Polymorphism, etc.)

Variation Attributes

Genetic Locus

Genetic Loci

Bio Sequence

Polypeptide

genotype $\leftrightarrow$ phenotype

Phenotype

Entry Point: Genome

Encapsulating Obj.

Bubbled-up Obj.

Non-locus specific data

Genetic Loci

Non-locus specific data
The HL7 RCRIM CT Laboratory Model- The Pharmacogenomics Extension

Utilizes the Clinical Genomics CMET

Clinical Trial

Enrolled Subject

Specimen

Pharmacogenomics Test

Consent to Genotype

Genetic Lab
The CG V3 Query Model: Query by Parameter

Starting point with query identifiers and attributes

Phenotype parameters

participants parameters

GeneticLoci parameters

Miscellaneous parameters

GeneticLocus parameters
V2 Implementation Guides

- A v2 laboratory message implementation guide for genetic testing result
  - A message from the genetic lab to the EHR
  - An approved informative spec

- It is modeled after the HL7 Version 2.5.1 Implementation Guide: Orders And Observations; Interoperable Laboratory Result Reporting To EHR (US Realm), Release 1

- Is used in a pilot of information exchange between Partners Healthcare and Intermountain Health Care
v2 Updates for May 2012

Updates prepared by Mollie Ullman-Cullere

- Genetic Variation – Release 2
  - Extends Lab Reporting IG HL7 2.5.1
  - Scope: Mutations, interpretations, and references for variations found within a gene

- Cytogenetics – Release 1
  - Extends Lab Reporting IG HL7 2.5.1
  - Scope: Description: Structures cytogenetic findings in accordance with adopted standards (ISCN nomenclature) while supporting narrative reporting of interpretation
  - Pilots: Looking for pilot organization(s)
The v2 Message Structure

Genetic analysis master panel (OBR)

Genetic Analysis Summary Panel (OBR)
(i.e. Case Definition in OBX’s)
including Medication or Disease Assessed,
Genomic Source Class, Analysis Report, and optional
Overall Interpretation

Genetic Analysis Discrete Result Panel (OBR)

DNA Analysis Discrete Sequence Variation Panel (OBR)
(i.e. Findings in OBX’s)
including Reference Sequence Identifiers, DNA Sequence Variation, Genomic Source Class, and optional Allele Name and Sequence Variation Interpretation
V2 Sample Message

- **OBR|1||PM-08-J00094^HPCGG-LMM^2.16.840.1.113883.3.167.1^ISO|Im_DCM-pnlB_L^Dilated Cardiomyopathy Panel B (5 genes)^99LMM-ORDER-TEST-ID||20080702000000|20080702100909||234567891^Pump^Patrick^^^^^NPI^L|20080703000000||F||00000009^Cardiovascular^99HPCGG-GVIE-INDICATION^^^^^Clinical Diagnosis and Family History of DCM|&Geneticist&Gene&&&NPI^^^^^HPCGG-LMM&2.16.840.1.113883.3.167.1&ISO|||55233-1^Genetic analysis master panel ^LN

- **SPM|1||119273009&Peripheral blood&SNM3&0707Intl&Blood, Peripheral|20080702000000

- **OBR|2||PM-08-J00094-1^HPCGG-LMM^2.16.840.1.113883.3.167.1^ISO|55232-3^Genetic analysis summary panel^LN||20080702000000|||20080703000000||F||^PM-08-J00094&HPCGG-LMM&2.16.840.1.113883.3.167.1&ISO

- **OBX|1|CWE|51967-8^Genetic disease assessed^LN||399020009^DCM-Dilated Cardiomyopathy^SNM3^^^0707Intl||F|20080702100909||Laboratory for Molecular Medicine^L^22D1005307^^^CLIA&2.16.840.1.113883.4.7&ISO|1000 Laboratory Lane^Ste. 123^Cambridge^MA^99999^USA^B
CDA IG: Genetic Testing Report (GTR)

- Define an implementation guide for a genetic testing report that is both human readable and machine-processable
  - Target at all types of GTR producers, e.g., genetic labs, clin. geneticists
  - Readable content is larger in scope
  - E.g., detailed description of the tests performed along with references
  - Machine-processable should be limited, e.g., exclude raw data

- Ballot a Universal IG; then derive specific types of GTR:
  - Healthcare & Research
  - Realm-specific guides
  - Omic-specific guides

- Developed using the MDHT open source tool (OHT)
GTR - Design Principles

- Follow existing report formats commonly used in healthcare & research

- Emphasize interpretations & recommendations

- Provide general background information on tests performed

- Reference HL7 Clinical Genomics instances (e.g., v3 or v2 GeneticVariation and Pedigree) as the place holders of full-blown raw genomic data and fully-structured family history data

- Utilize patterns of ‘genotype-phenotype’ associations in the HL7 v3 Clinical Genomics Domain
  - Implement them as ‘clinical genomic statement’ entry-level templates (see next slide), enabling meaningful use of the data
The Clinical Genomic Statement

- An abstract Clinical Genomic Statement (CGS) template that
  - Has at its core a genomic observation (e.g., a DNA sequence variation)
  - If it’s a reportable finding, then it should be associated with indications and interpretations, specimen and genomic source class
  - The major finding can be associated with associated observation (e.g., amino acid change)
  - Optionally, performers may be specified (overriding header performers)

- The CGS abstract template is instantiated by specialized CGS’s, e.g., for genetic variations or cytogenetics
Detailed description can be found in the GTR presentation
CDA GTR Ballot Status

- Balloted as DSTU and passed in Oct. 10 & Sep. 12
- Under reconciliation of ballot comments
- Main issues:
  - Vocabulary:
    - Mainly - Section codes
    - Should we mandate specific terminologies in structured data?
  - Layout:
    - Semantics – compare to recommended layouts in the literature
    - Syntactic – work closely with MDHT developers to adhere to SDWG guidelines
  - Sections specific to every type of genetic test (resolved)
  - Section and Entry level template ids registration (when layout agreed)
  - Suggestion to add drug safety template (considered for future use)
  - Tissue typing / marrow donor section (work in progress)
Collaboration with Anatomic Pathology

- Development:
  - How do we share templates?
  - MDHT SVN could be an option

- Harmonization:
  - One consolidated document template or two interrelated and harmonized docs?
  - Workflows could dictate the way documents shape up…
  - Could we share specimen templates?
  - Could we harmonize summary / diagnosis templates?
  - Could we agree on a clinical statement approach where all pieces are tied together in one coherent structure?

- Comments on the APSR sample:
  - Is problem == reason for AP procedure?
  - Why “Left upper outer quadrant breast palpable mass” is considered illness?
  - APSR sample has CDA Validation issues
Alignment Among the Various Specs

- v3 specs and CDA are all based on the RIM
  - CDA GTR-IG will be based on CDA R3
  - Depending on the “right side” of R3, if it allows RIM-based domain models, then alignment is trivial

- v3-v2 alignment:
  - Proposal: represent semantics with v3 and implement it in various ways, one of which is v2; develop an “v2 ITS” for the v3 models
  - See proposal made by Amnon in a separate presentation (click here to see that presentation)
Utilizations in HL7

- **Clinical Trials:**
  HL7 RCRIM Work Group (clinical trials specs) utilized the CG DSTU model (Genetic Locus) in their Pharmacogenomics message, which was an extension of the CTLab message (an approved but expired DSTU)

- **Laboratory:**
  The Lab Work Group might utilize a constrained version of the Genetic Variation model in their next release if the Lab Result message
Selected Implementation

- **v2**
  - Exchange of genetic testing results between Intermountain and Harvard

- **v3**
  - The Family History spec is used in Mass General Hospital
  - Expanding to other family history applications including the US Surgeon General Family History tool
  - The Genetic Variation model is used in Hypergenes (a European project on essential hypertension, [http://www.hypergenes.eu/](http://www.hypergenes.eu/))
  - The Pedigree and Genetic Variation models are used in Italy, the Rizzoli institute in Bologna

- **CDA**
  - GTR has been used in uHealth – a PHR/EHR system in Korea
Domain Analysis Models

Work in progress…

- Clinical (Gen)omics DAM
- Gene Expression DAM
- Clinical Sequencing DAM
- Specimen
  - Specimen DAM
  - Specimen Unique ID
  - Specimen CMET Release 2
Oomics in the LS DAM - Experiment (outdated)
Gene Expression Topic (outdated)

- **Domain Analysis Model (DAM)**
  - Passed informative ballot
  - Based on several models for gene expression data along with extensions
Clinical Sequencing Analysis Model

Update prepared by Mollie and Daryl (May 2012)

- Extends ONC Personalized Healthcare Use Case (2008) for support of clinical sequencing and current understanding of the field

- Serve as
  - Roadmap for further standards development/extension
  - Develop use cases and requirements for related projects (e.g. Specimen DAM and Specimen CMET Release 2)
Different compatible standards may need to be used at different parts in the workflow (e.g. lab order vs. result reporting). Therefore, starting with workflow definition and key data needs is important (e.g. clinical indication for ordering the test).
These serve as potential integration points (e.g. dbSNP ID for links to dbSNP variant record).
Looking forward....
The rise of the 'narciss-ome'...

- Transformative paper in the Cell Journal
- Reviewed in Nature
  (http://www.nature.com/news/the-rise-of-the-narciss-ome-1.10240)

- iPOP – Integrative Personal Omics Profile
  - Our personal omics change over time!
  - Longitudinal examinations of genome, proteome, metabolome, autoantibodies, etc. of an individual (one of the authors)
  - Monitor healthy and disease states
  - Predict and act accordingly (the data predicted diabetes and diabetes was diagnosed; life style changes make it manageable!)
iPOP – Integrative Personal Omics Profile

Should be standardized and be part of the lifetime EHR!!
HL7 WG Health Check – Need to Improve!

- Active projects
- SWOT current
- 3 year plan current
- Mission and charter current
- Co-chair post-WGM survey participation
- Ballot presence
- Minutes posted since last WGM
- Last listserv activity
- Wiki presence
- WG conference calls schedules
- Steering division conference call participation
- Steering division co-chair (TSC representation) election participation
- WG rep at steering division WGM
- WG meetings at WGM scheduled
- WG has an approved DMP based on review of the updated template
Planning ahead

- January 2013 WGM (Phoenix)
  - Schedule (from Tuesday Q3 to Thursday Q2)
  - Joint meetings with:
    - AP
    - RCRIM
    - OO+AP+II

- Weekly conf. calls
  - Continue Tuesday’s 11EST

- Prepare to ballot GTR, Family History IG for US Realm, Sequencing DAM in January 2013
Summary

- Small group coping with
  - Various HL7 formats: v3, v2 and CDA
  - Clinical & Research environments

- Developing a DAM and component models (CMETs) to be used in other HL7 domains
  - Genetic Variation
  - Gene Expression

- CDA Genetic Testing Report (GTR)
  - Bridge from raw data to human readable reports and bubbled-up data
  - Model-driven development of standards (use of MDHT CDA Editor)
A new member? Here is how to participate:

- Through our mailing list (open to nonmembers as well)
- Provide comments to specs
  - Formally through the balloting process
  - Informally through the list or by corresponding with spec editors
- Weekly conference calls
- Project specific conf. calls:
  - Clinical sequencing
  - CG & GE DAMs
  - Specimen
The End

• Thank you for your attention… 😊

• Questions? Contact Amnon at shabo@il.ibm.com

• Comments of general interest should be posted to the CG mailing list at clingenomics@lists.hl7.org