FHIR Genomics Use Case: Reporting HLA Genotyping

Consensus-Sequence-Block & Observation-genetics-Sequence Cardinality

Bob Milius, PhD
Bioinformatics
NMDP/Be The Match

NATIONAL MARROW DONOR PROGRAM®

BE THE MATCH®

Outline

- Background
  - MIRING, & HML
  - FHIR Profiles Created for HLA
    - Diagnostic Report > Profile for HLA Genotyping Results
    - Sequence > Profile for Consensus Sequence Blocks
- Proposal #1
  - Delete Sequence > Profile for Consensus Sequence Blocks
- Proposal #2
  - Change cardinality of Observation-genetics-Sequence from 0..1 to 0..*

Human Immunology 76 (2015) 954–962

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Minimum information for reporting next generation sequence genotyping (MIRING): Guidelines for reporting HLA and KIR genotyping via next generation sequencing

Steven J. Mack [a,*], Robert P. Milius [b], Benjamin D. Gifford [c], Jürgen Sauter [d], Jan Hofmann [d], Kazutoyo Osoegawa [e], James Robinson [f,g], Mathijs Groeneweg [h], Gregory S. Turenchalk [i], Alex Adai [i], Cherie Holcomb [j], Erik H. Rozemuller [k], Maarten T. Penning [k], Michael L. Heuer [b], Chunlin Wang [l], Marc L. Salit [m,n], Alexander H. Schmidt [d], Peter R. Parham [o], Carlheinz Müller [p], Tim Hague [q], Gottfried Fischer [r], Marcelo Fernandez-Viña [e], Jill A. Hollenbach [s], Paul J. Norman [o], Martin Maiers [b]

[a] Children's Hospital Oakland Research Institute, Oakland, CA, USA
[b] National Marrow Donor Program, Minneapolis, MN, USA
[c] One Lambda, Thermo Fisher Scientific, Brown Deer, WI, USA
[d] DKMS German Bone Marrow Donor Center, Tübingen, Germany
[e] Department of Pathology, Stanford University, Stanford, CA, USA
[f] Anthony Nolan Research Institute, Royal Free Hospital, London, UK
[g] University College London Cancer Institute, University College London, London, UK
[h] Maastricht University Medical Centre, Maastricht, The Netherlands
[i] Bioinformatics, Roche Sequencing, Pleasanton, CA, USA
[j] Roche Molecular Systems, Pleasanton, CA, USA
[k] GenDX, Utrecht, The Netherlands
[l] Stanford Genome Technology Center, Stanford University, Stanford, CA, USA
[m] National Institute of Standards and Technology, Stanford, CA, USA
[n] Department of Bioengineering, Stanford University, Stanford, CA, USA
[o] Department of Structural Biology, Stanford University, Stanford, CA, USA
[p] Zentrales Knochenmarkspender-Register Deutschland, Ulm, Germany
[q] Omixon, Budapest, Hungary
[r] Medical University of Vienna, Vienna, Austria
[s] Department of Neurology, University of California, San Francisco, CA, USA

Human Immunology 76 (2015) 963–974

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Histoimmunogenetics Markup Language 1.0: Reporting next generation sequencing-based HLA and KIR genotyping

Robert P. Milius [a,*], Michael Heuer [a], Daniel Valiga [a], Kathryn J. Doroschak [a], Caleb J. Kennedy [a], Yung-Tsi Bolon [a], Joel Schneider [a], Jane Pollack [a], Hwa Ran Kim [b], Nezih Cereb [b], Jill A. Hollenbach [c], Steven J. Mack [d], Martin Maiers [a]

[a] National Marrow Donor Program, MN, USA
[b] HistoGenetics LLC, USA
[c] University of California – San Francisco, USA
[d] Children's Hospital & Research Center Oakland, Oakland, CA, USA

ARTICLE INFO

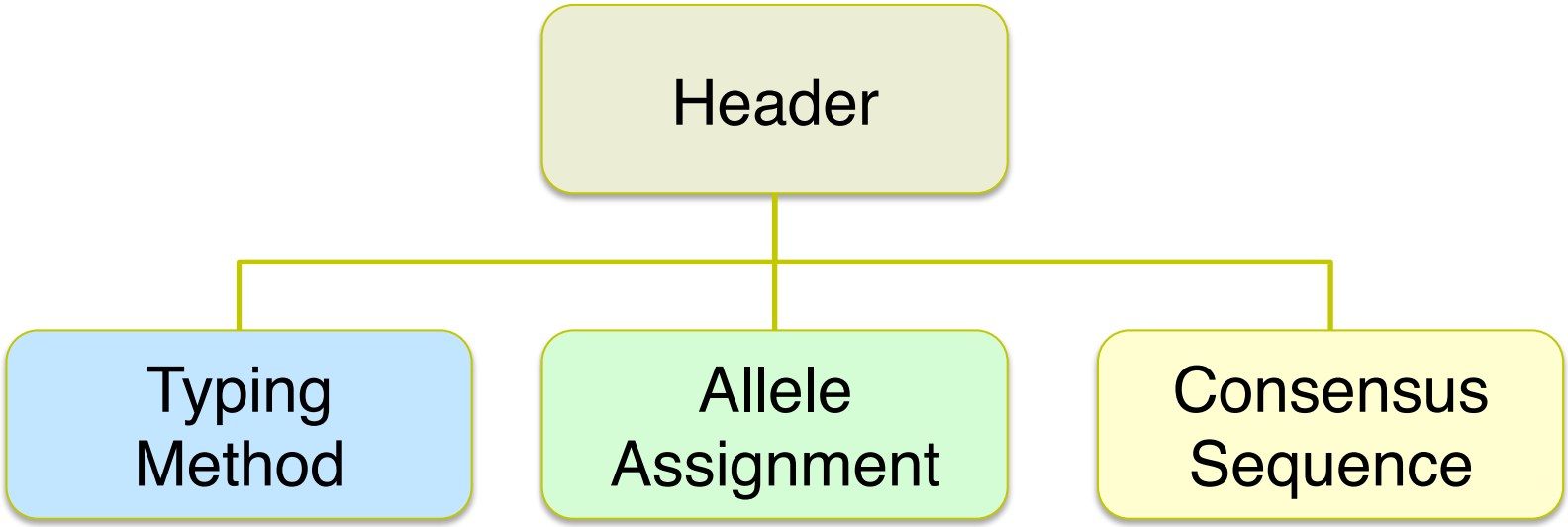
Article history:
Received 31 January 2015
Revised 30 May 2015
Accepted 6 August 2015
Available online 28 August 2015

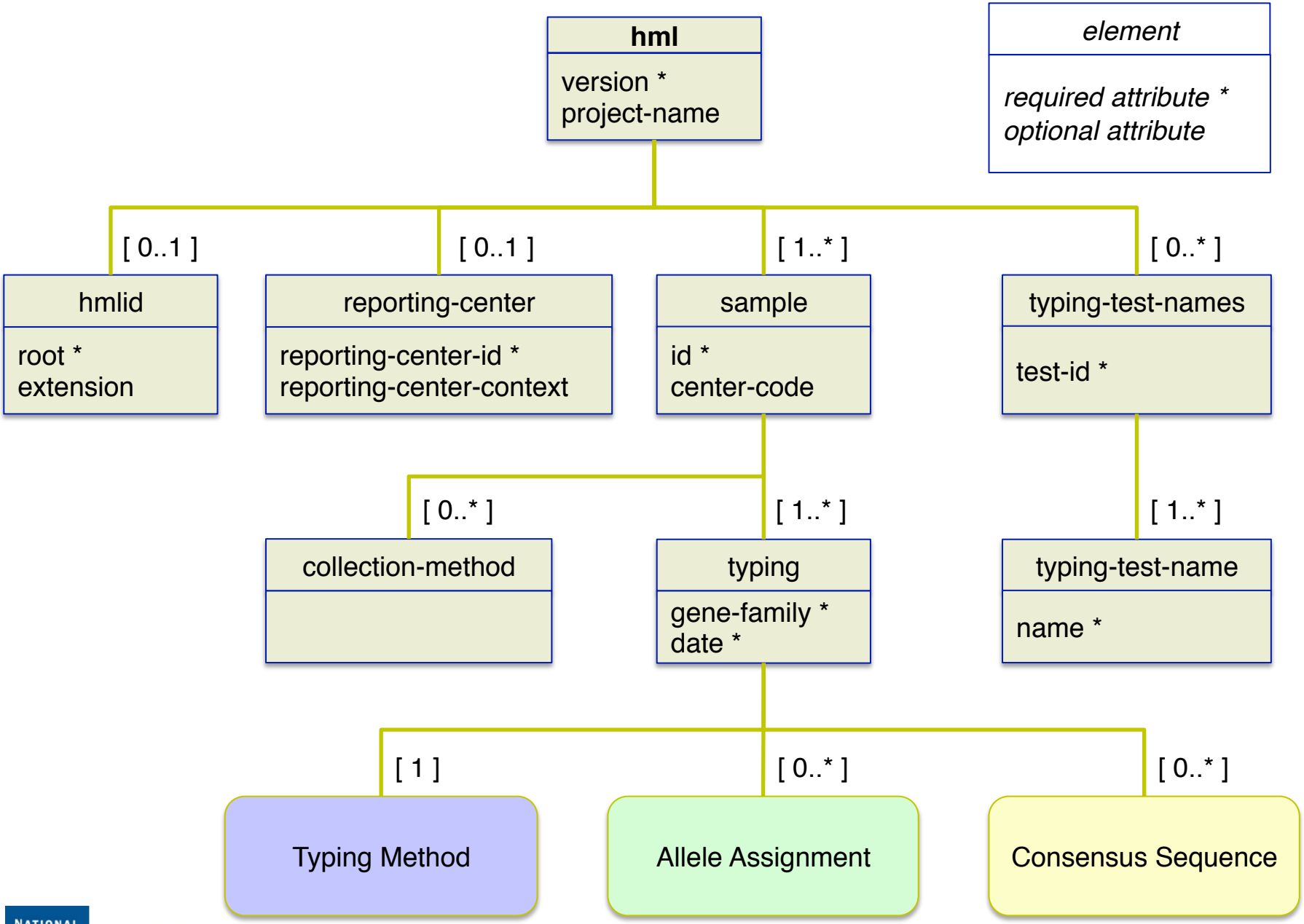
Keywords:
HML
NGS
HLA
KIR
MIRING
Data standards
Genotyping

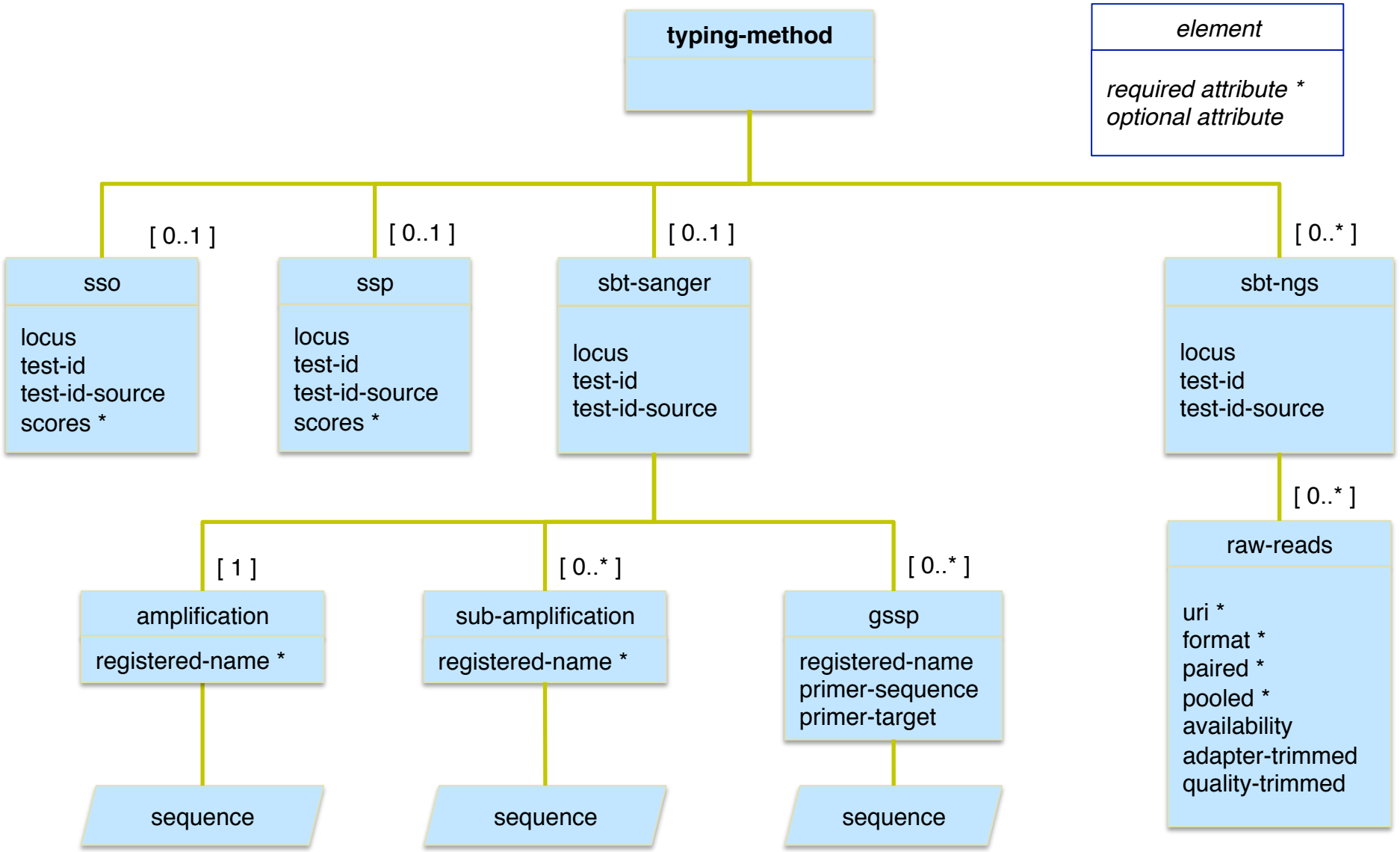
ABSTRACT

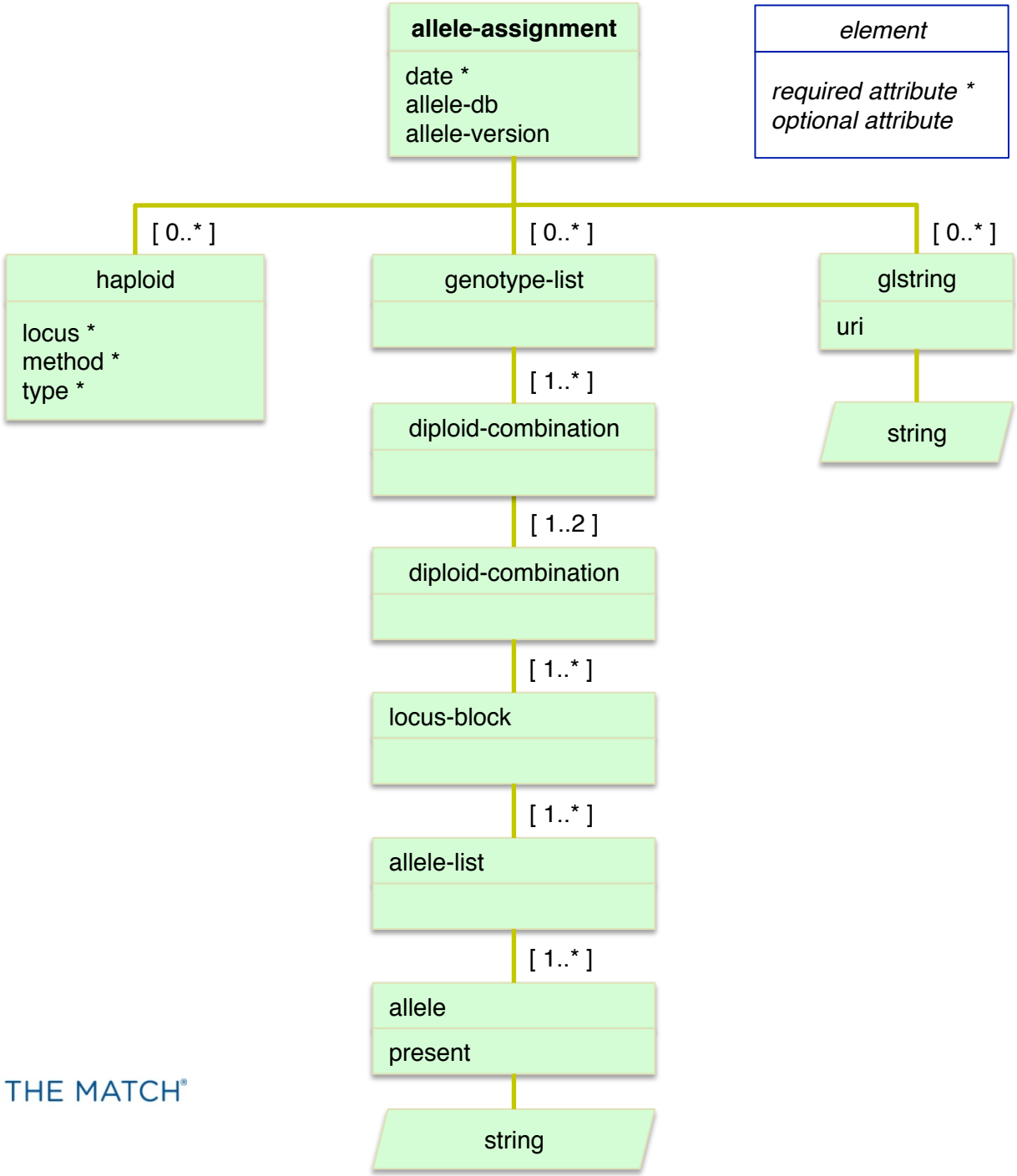
We present an electronic format for exchanging data for HLA and KIR genotyping with extensions for next-generation sequencing (NGS). This format addresses NGS data exchange by refining the Histoimmunogenetics Markup Language (HML) to conform to the proposed Minimum Information for Reporting Immunogenomic NGS Genotyping (MIRING) reporting guidelines (miring.immunogenomics.org). Our refinements of HML include two major additions. First, NGS is supported by new XML structures to capture additional NGS data and metadata required to produce a genotyping result, including analysis-dependent (dynamic) and method-dependent (static) components. A full genotype, consensus sequence, and the surrounding metadata are included directly, while the raw sequence reads and platform documentation are externally referenced. Second, genotype ambiguity is fully represented by integrating Genotype List Strings, which use a hierarchical set of delimiters to represent allele and genotype ambiguity in a complete and accurate fashion. HML also continues to enable the transmission of legacy methods (e.g. site-specific oligonucleotide, sequence-specific priming, and Sequence Based Typing (SBT)), adding features such as allowing multiple group-specific sequencing primers, and fully leveraging techniques that combine multiple methods to obtain a single result, such as SBT integrated with NGS.

HML 1.0


Header
Typing Method
Allele Assignment
Consensus Sequence

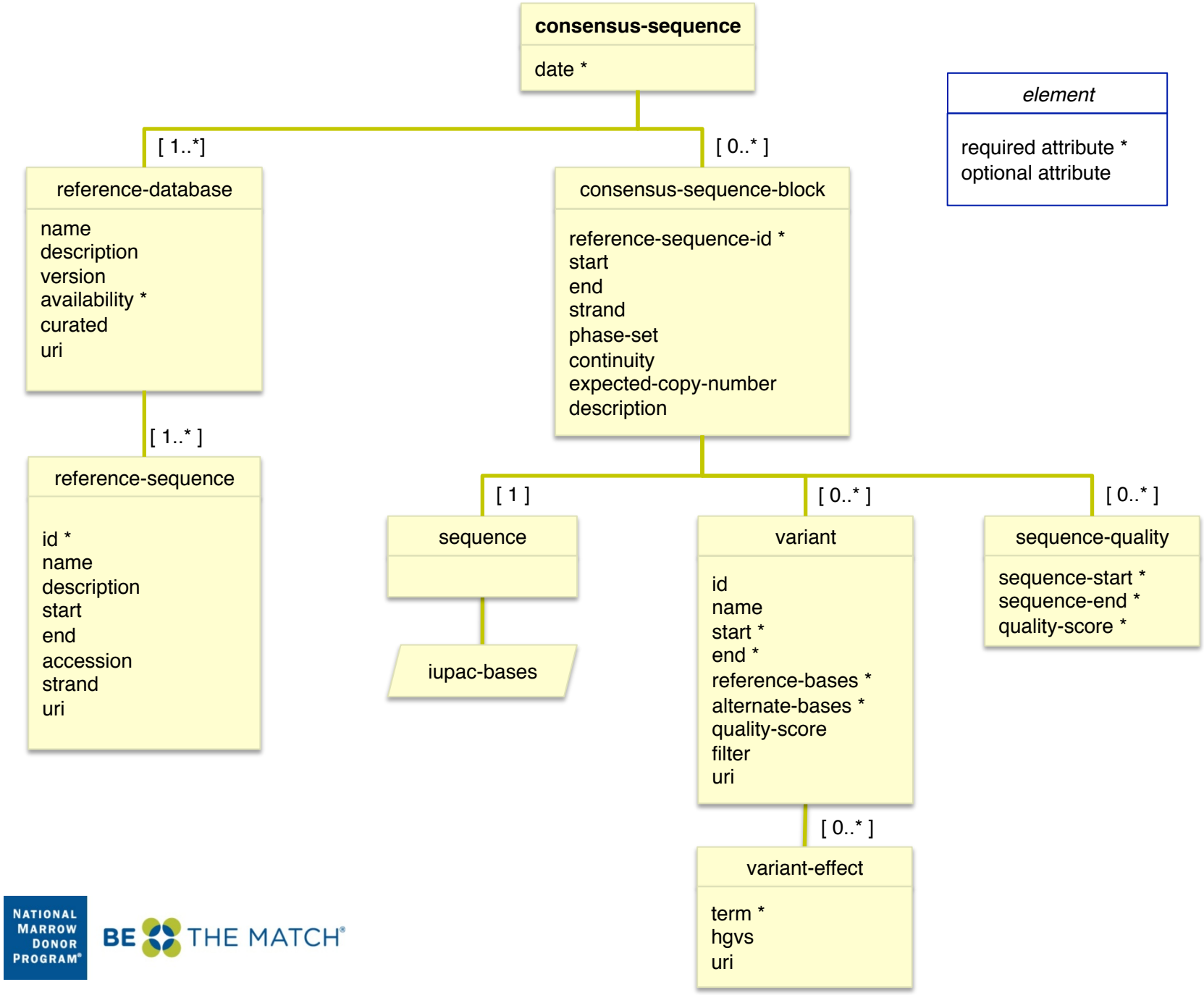

hml
version *
project-name
element
required attribute *
optional attribute
[0..1]
[0..1]
[1..*]
[0..*]
hmlid
root *
extension
reporting-center
reporting-center-id *
reporting-center-context
sample
id *
center-code
typing-test-names
test-id *
[0..*]
[1..*]
[1..*]
collection-method
typing
gene-family *
date *
typing-test-name
name *
[1]
[0..*]
[0..*]
Typing Method
Allele Assignment
Consensus Sequence


typing-method
element
required attribute *
optional attribute
[0..1]
[0..1]
[0..1]
[0..*]
sso
locus
test-id
test-id-source
scores *
ssp
locus
test-id
test-id-source
scores *
sbt-sanger
locus
test-id
test-id-source
sbt-ngs
locus
test-id
test-id-source
[0..*]
raw-reads
uri *
format *
paired *
pooled *
availability
adapter-trimmed
quality-trimmed
[1]
[0..*]
[0..*]
amplification
registered-name *
sub-amplification
registered-name *
gssp
registered-name
primer-sequence
primer-target
sequence
sequence
sequence



NATIONAL MARROW DONOR PROGRAM®

BE THE MATCH®



NATIONAL MARROW DONOR PROGRAM®

BE THE MATCH®

Two FHIR Profiles were created for the HLA reporting use case

- Diagnostic Report > Profile for HLA Genotyping Results
  - http://hl7-fhir.github.io/diagnosticreport-hla-results.html

- Sequence > Profile for Consensus Sequence Blocks
  - http://hl7-fhir.github.io/sequence-consensus-sequence-block.html

HLA Profiles
Diagnostic Report > Profile for HLA Genotyping Results

- Diagnostic Report > Profile for HLA Genotyping Results
  - http://hl7-fhir.github.io/diagnosticreport-hla-results.html

- hla-genotyping-resultsAlleleDatabase
  - **Allele Database** : Allele Database.
- hla-genotyping-resultsAlleleAsignmentDate
  - **AlleleAsignment-data** : AlleleAsignment-data.
- hla-genotyping-resultsGlstring
  - **glstring** : glstring.
- hla-genotyping-resultsHaploid
  - **haploid** : haploid.
- hla-genotyping-resultsConsensusSequenceBlock
  - **ConsensusSequenceBlock** : ConsensusSequenceBlock.
- hla-genotyping-resultsMethod
  - **The platform, methodology and software applied at the time of the genotyping** : The platform, methodology and software applied at the time of the genotyping.

HLA Profiles
Sequence > Profile for Consensus Sequence Blocks

- http://hl7-fhir.github.io/sequence-consensus-sequence-block.html

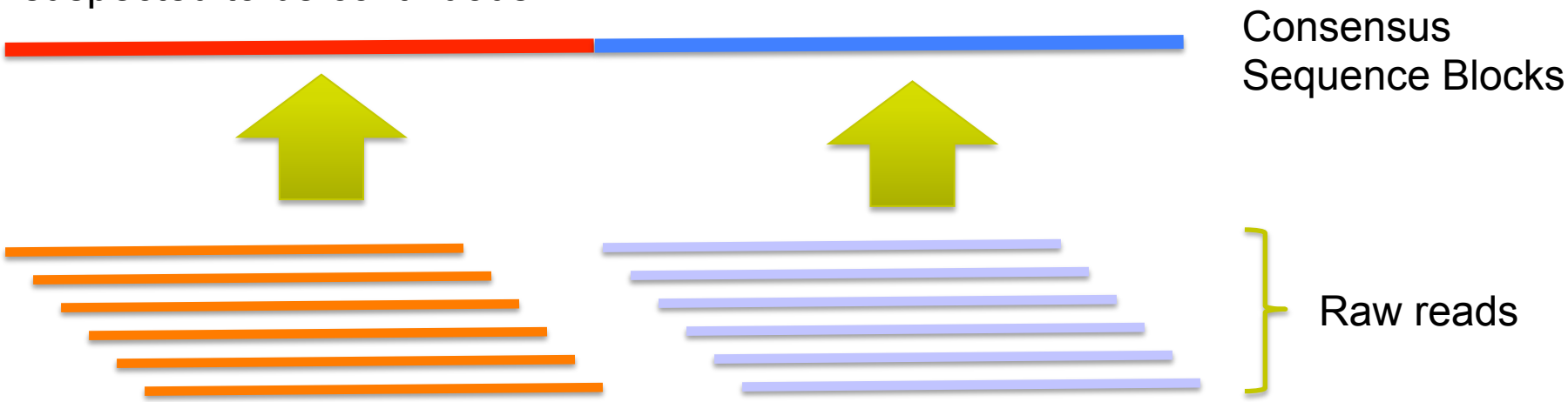
- sequence-consensus-sequence-blockExpectedCopyNumber
  - **Expected Copy Number** : To indicated how many copies of the sequence block was expected.
  - May be satisfied with observation-geneticsAllelicState

- sequence-consensus-sequence-blockStrand
  - **Strand** : Strand.
  - May be satisfied with Sequence.referenceSeq.strand

- sequence-consensus-sequence-blockContinuity
  - **Continuity** : A continuity flag to indicate whether a gap exists between blocks.
  - May be calculated with coordinates to a reference and calculated indels

Sequence Continuity

Two consensus sequence blocks (contigs) that are phased but without a bridging sequence between them. We use a “continuity” attribute in HML to indicated if two sequence blocks are separated by a gap or if they are suspected to be continuous.



We can calculate this using sequence coordinates relative to a reference (taking into account any indels), so perhaps we don’t need this.

First Proposal

- Remove a Profile
  - Sequence > Profile for Consensus Sequence Blocks

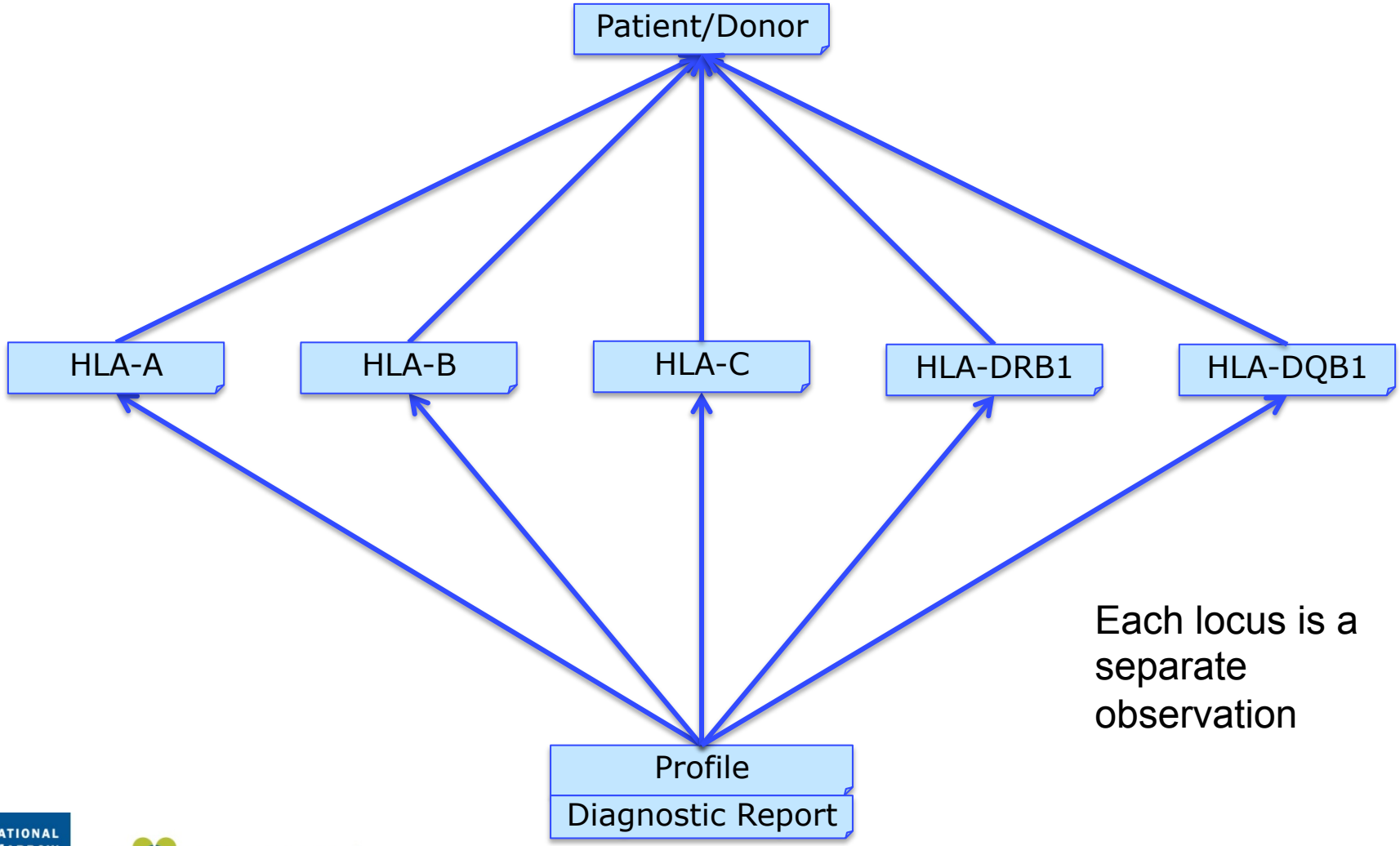
- Why?
  - Almost all the extensions found in this profile can now be met using existing elements in Sequence and Observation for Genetics,
  - The one remaining extension (continuity) may be calculated, and it doesn't make sense and in its current structure (continuity describes the relationship between two sequences, not an attribute of a single sequence).

Observation-genetics-Sequence Cardinality

- Observation-genetics-Sequence
  currently 0..1

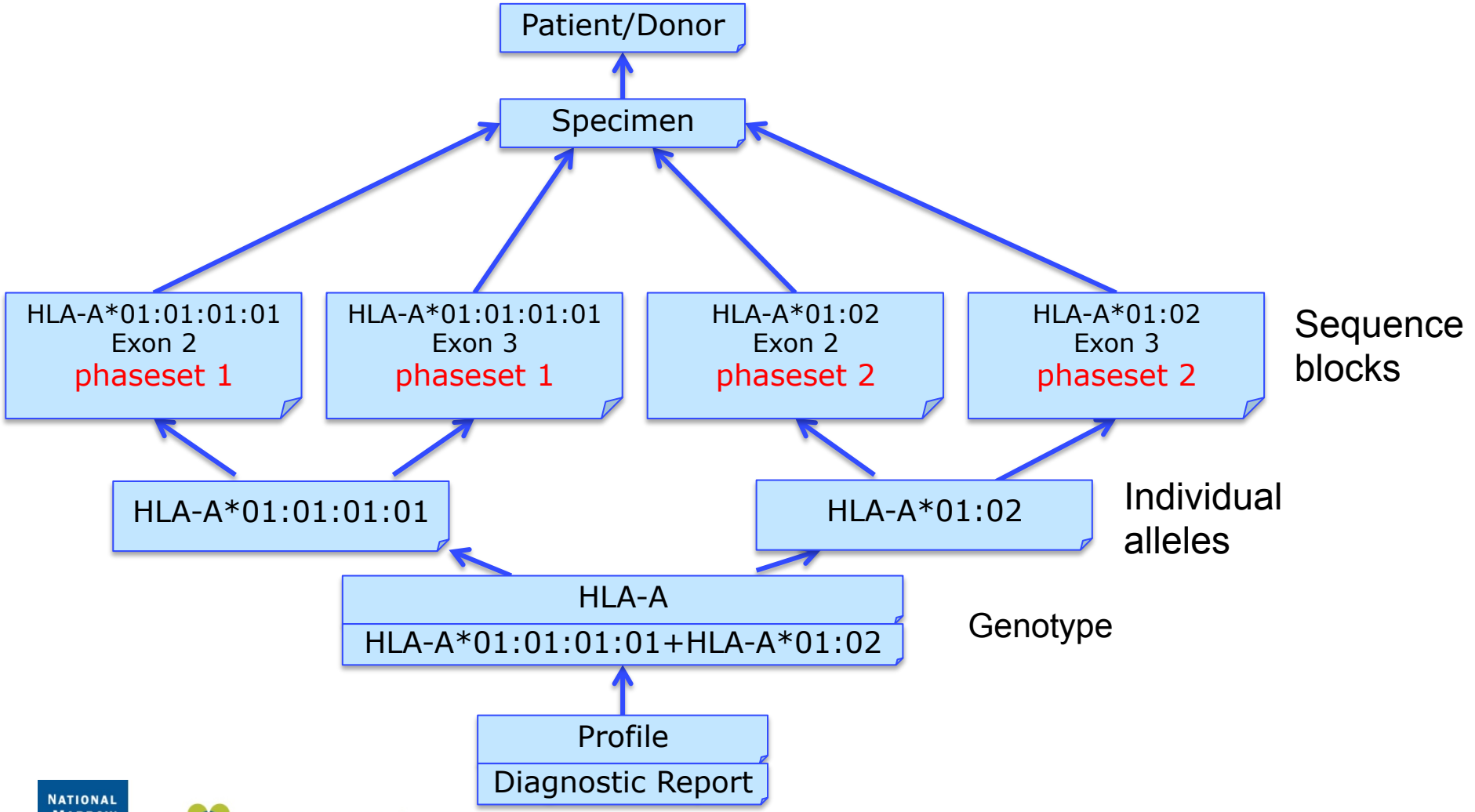
- Why not 0..* ?

Need to report genotypes of a panel of HLA loci in a single report

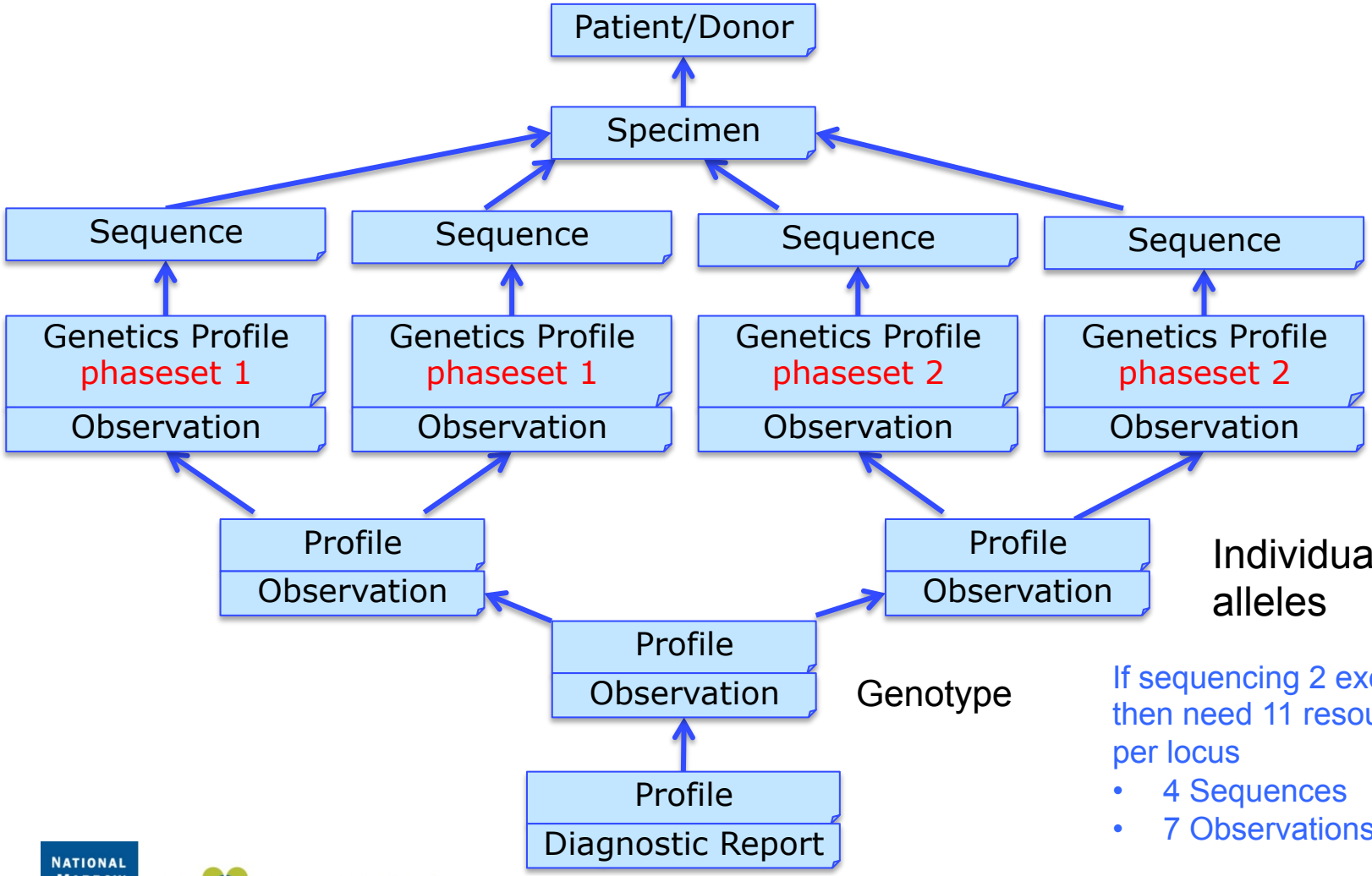


Each locus is a separate observation

Breaking down a genotype for a single locus, e.g. HLA-A



Observation-genetics-Sequence 0..1



Sequence blocks

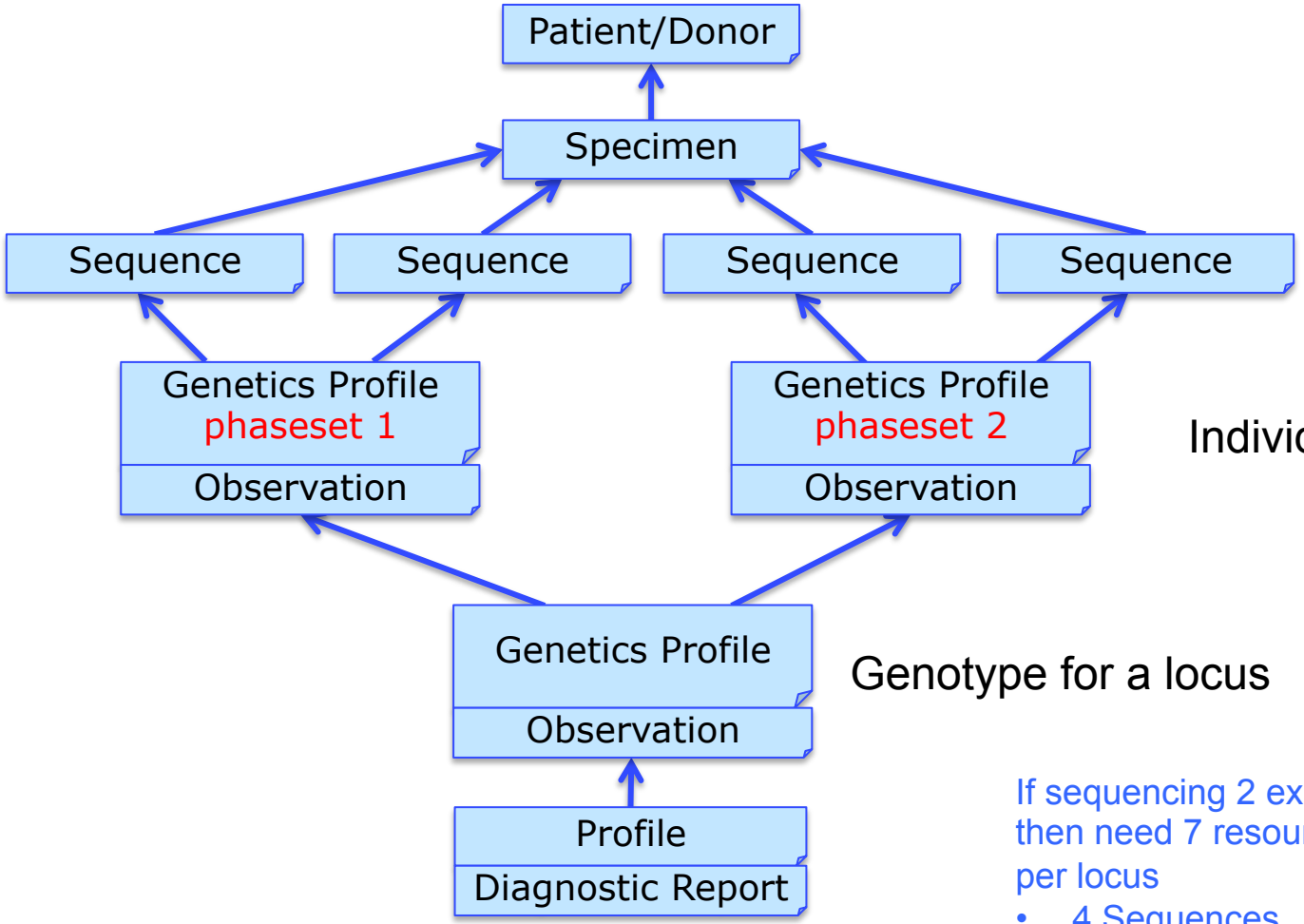
Individual alleles

Genotype

If sequencing 2 exons per locus, then need 11 resource instances per locus

- 4 Sequences
- 7 Observations

Reducing the number of Observations needed by changing cardinality of Observation-genetics-Sequence to 0..*



Individual alleles

Genotype for a locus

If sequencing 2 exons per locus, then need 7 resource instances per locus

- 4 Sequences
- 3 Observations

Summary: Two Proposals

1. Remove a Profile
   - Sequence > Profile for Consensus Sequence Blocks

2. Change cardinality of Observation-genetics-Sequence from 0..1 to 0..*