**Overall description:**

Intron positions and intron sequences from lrg files too look for things like cryptic splice sites.

Can pull out all introns to look for cryptic splice sites in the future

Needed because genomic information is being returned and introns are not being looked at and investigated.

**What we need to do today:**

* Look at LRG
* See about how data is structured
* And think about what we might need to do
* Inputs needed
* Outputs generated
* Why we want these things

**Notes about what we might need:**

Pending approval warning

Make sure that the LRG file is from the correct build – display the sequence

Display reference sequence ID (NM number)

**The gene we are using to test with:**

*PMP22*:

* Involved in inherited peripheral neuropathies
* 5 exons

The LRG website uses RELAX NG schema language for XML. The key features of RELAX NG are that it:

* is simple
* is easy to learn
* has both an XML syntax and a compact non-XML syntax
* does not change the information set of an XML document
* supports XML namespaces
* treats attributes uniformly with elements so far as possible
* has unrestricted support for unordered content
* has unrestricted support for mixed content
* has a solid theoretical basis
* can partner with a separate datatyping language (such W3C XML Schema Datatypes)

**Programs to help make XML more readable:**

Dreamweaver

G edit

Notepad ++

**Information about the XML schemer in LRG:**

<ftp://ftp.ebi.ac.uk/pub/databases/lrgex/docs/LRG_XML_schema_documentation_1_9.pdf>