

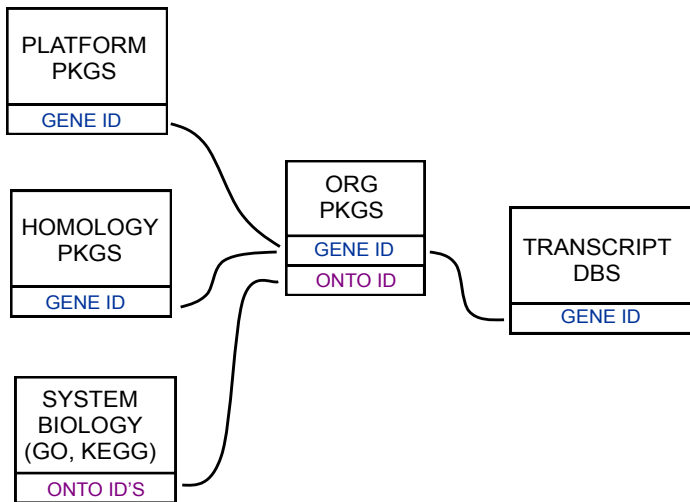
Annotation Packages

Marc Carlson

Fred Hutchinson Cancer Research Center

July 28th 2010

Bioconductor Annotation Packages



The newest addition are *Bioconductor* Transcript DBs

The GenomicFeatures package

- ▶ Transcript information stored in SQLite databases
- ▶ These databases attempt to represent the relational nature of splicing data correctly
- ▶ Transcript information exposed to user via the *GenomicRanges* infrastructure
- ▶ Parsers are available to construct these databases from biomaRt or the UCSC genome browser. Also there is a generic parser for custom jobs.

Some transcript sources for *GenomicFeatures*

Constructors

makeTranscriptDbFromBiomart, makeTranscriptDbFromUCSC

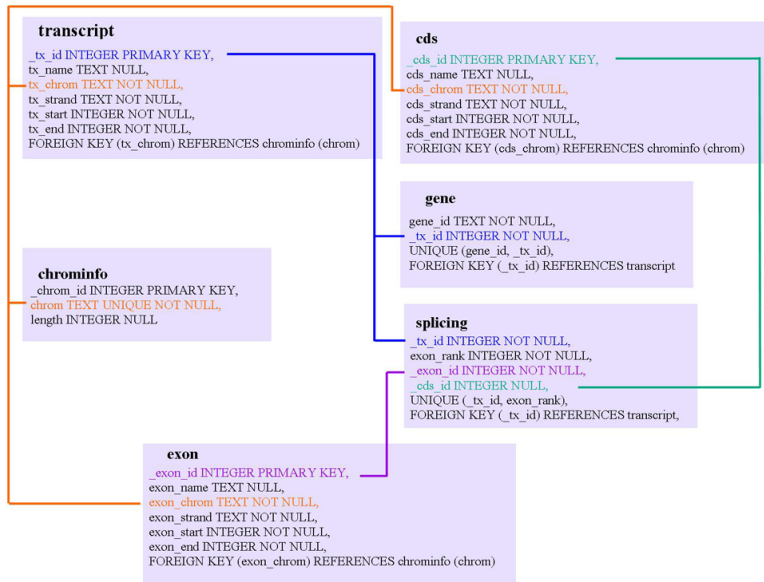
```
> library(GenomicFeatures)
> nrow(supportedUCSCtables())
```

```
[1] 24
```

```
> head(supportedUCSCtables(), 10)
```

	track	subtrack
knownGene	UCSC Genes	<NA>
knownGeneOld3	Old UCSC Genes	<NA>
wgEncodeGencodeManualRel12	Gencode Genes	Genecode Manual
wgEncodeGencodeAutoRel12	Gencode Genes	Genecode Auto
wgEncodeGencodePolyaRel12	Gencode Genes	Genecode PolyA
ccdsGene	Consensus CDS	<NA>
refGene	RefSeq Genes	<NA>
xenoRefGene	Other RefSeq	<NA>
vegaGene	Vega Genes	Vega Protein Genes
vegaPseudoGene	Vega Genes	Vega Pseudogenes

The *TranscriptDb* schema



Future *Bioconductor* Annotation Packages?

