# Stem Cell Engineering Informatics in 2015

Status of LIMS2, HTGT and WGE systems

David Parry-Smith
17 March 2015

# Talk Outline

- What are we trying to achieve with our systems?
- Overall landscape of systems
- HTGT – legacy but still used
- LIMS2 – the current LIMS still actively developed
- WGE – a research tool for CRISPR genome editing
- Documentation
- Future plans
- People

# What are we aiming to achieve?

- Tracking systems for wells that are manipulated in the lab
- Tools for reporting over those wells
- Tools for engineering designs
- Tools for designing, tracking and reporting CRISPR experiments

# The Landscape

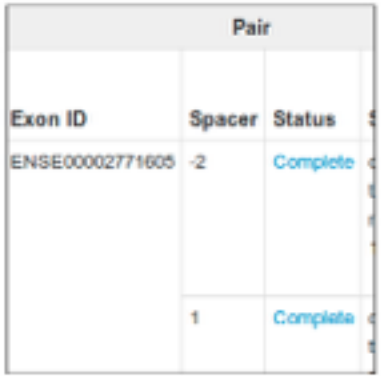
URL: www.sanger.ac.uk/htgt

## High Throughput Gene Targeting

### Team 87

HTGT2

LIMS2




WGE

# HTGT

- HTGT – High-Throughput Gene Targeting
  - The ‘original’ LIMS for Team87
  - Perl/Catalyst/DBI/DBIx::Class/Oracle
    - Catalyst 5.90042
  - Was running on Etch (aka ancient version of Debian)
  - A ‘legacy’ system supporting Mouse (only)
  - Essentially relational ‘one well -> one parent’ model
    - A kind of informal tree structure

# HTGT2 – The migration

- Removal of systems support for Etch O/S (summer 2013)
  - drives effort to migrate to Ubuntu Precise (VMs)
  - The painstaking task of gathering all Perl modules required:
    - Local – multiple subversion repositories re-organised into handful of GIT repositories (using GIT submodules)
      - htgt-root
      - htgt-batch
      - htgt-app
    - CPAN modules for Perl 5.14.4 ...

# HTGT2 code repo

| | |
|---|---|
| Eng-Seq-Builder @ 29fef7f | Addition of deletion annotation and transcript id into genbank files ... |
| HTGT-QC-Common @ 196a... | Added z1a and z2a primers to eucomm-post-cre qc profile. Updated modules |
| JobRunner @ 540adad | add JobRunner & ruby |
| LIMS2-REST-Client @ 6e969a3 | add submodules Eng-Seq-Builder HTGT-QC-Common LIMS2-REST-Client |
| bin | Add devel directory to rsync to. |
| config | use scratch109 for QC while upgrade to 110 is in progress |
| data/mutant_sequences | Config and data dir changes |
| htgt_app @ 29ffe80 | updating commits |
| htgt_batch @ 84c24c4 | Merged in App and Batch modules |
| imits-perl-api @ 2d6bb63 | modify setup; add imits-perl-api |
| logs | initial changes for batch config files |
| logs_parser | Add csv clean-up job to job-runner; alter parse.rb for new piq job. |
| migration | Noted syntax change to CSV plugin; updated htgt_app |
| perl5 @ 5d48acd | updating commits |
| test | add test for new translate_first_reading_frame method added to HTGT::... |
| .gitignore | Added htgt-lims2 authentication key to config files |
| .gitmodules | modify setup; add imits-perl-api |
| README.md | Corrected file designation |
| setup.sh | Changes to bin directory and new setup.sh script to manage different ... |

# LIMS2

- Supports:
  - both Human and Mouse species
  - Gibson Designs
  - CRISPR/Cas9 support
- Runs on Ubuntu 10.04 (Lucid lynx)
  - Perl 5.10.1 (...old!)
  - Catalyst/DBIx::Class/PostgreSQL
    - Catalyst 5.90011
  - More use of JavaScript
  - ExtJS for flexible table widget
  - Genoverse genome browser

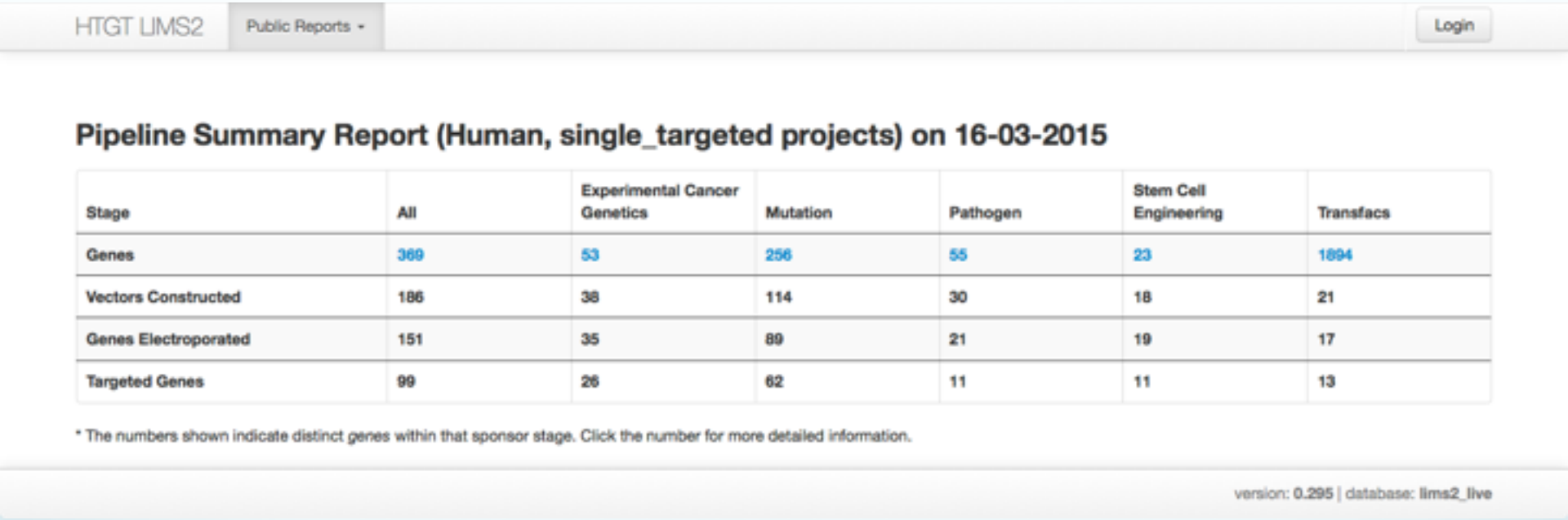
# LIMS2 - architecture

- A well can have ... multiple parent wells!
- Re-arrays, extensive parameter validation
- Access the DB through the model.
  - Thin controller/fat model.
- Shares QC system with HTGT2
  - For Farm3 submission
- Directed Acyclic Graph
- Graph Support is limited in DBIx::Class and ANSI SQL
  - 'WITH RECURSIVE' required in PostgreSQL to permit searching across the graph
    - No support in DBIx::Class ... yet
    - Required for fast reporting

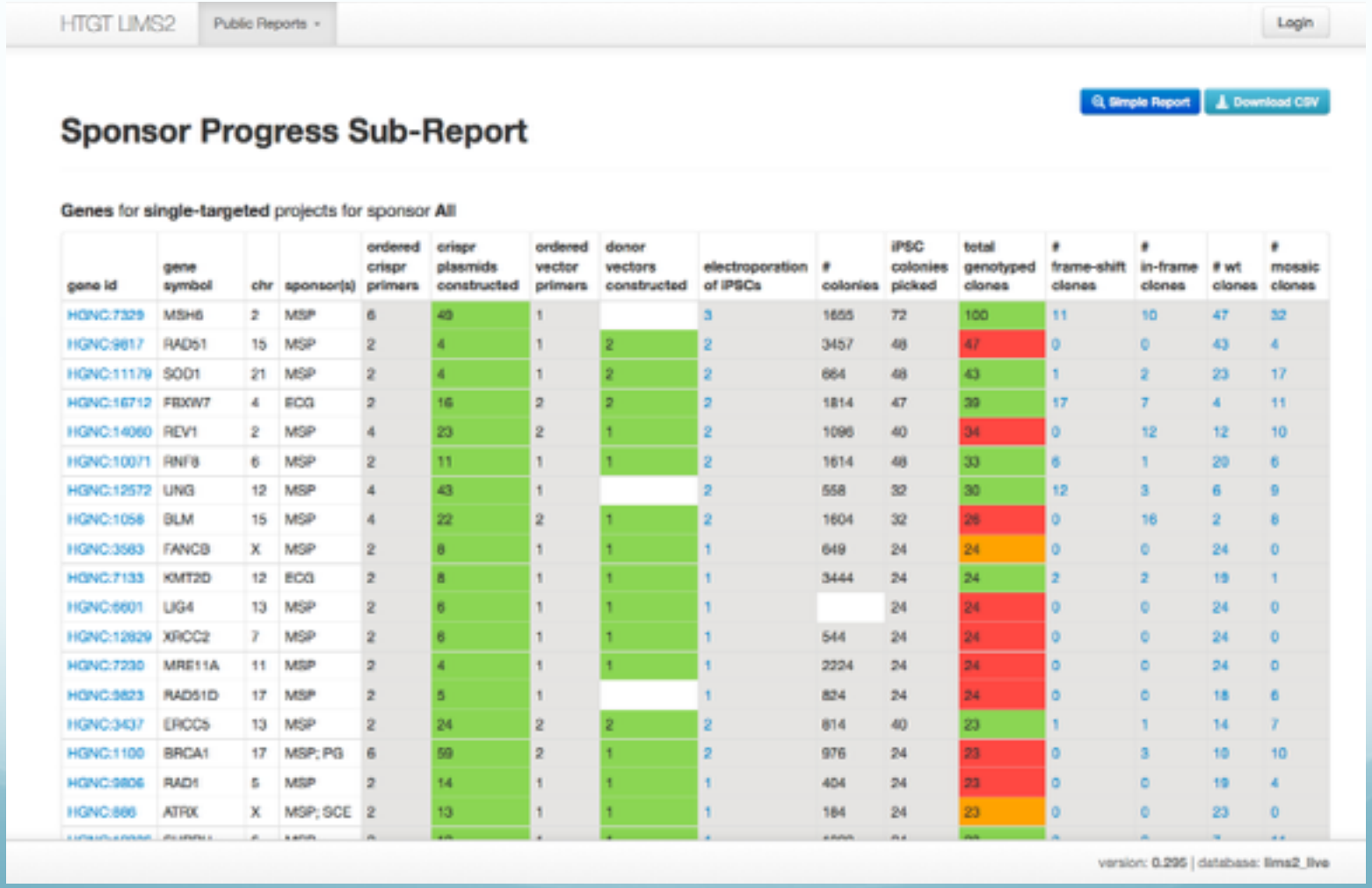
# Recent features

- Automatic genotyping/PCR/sequencing primer generation
- JavaScript trace viewer for Sequencing QC review
- CRISPR QC
- Bar coding check-in/check-out and tracking
  - Plate versioning required as a result

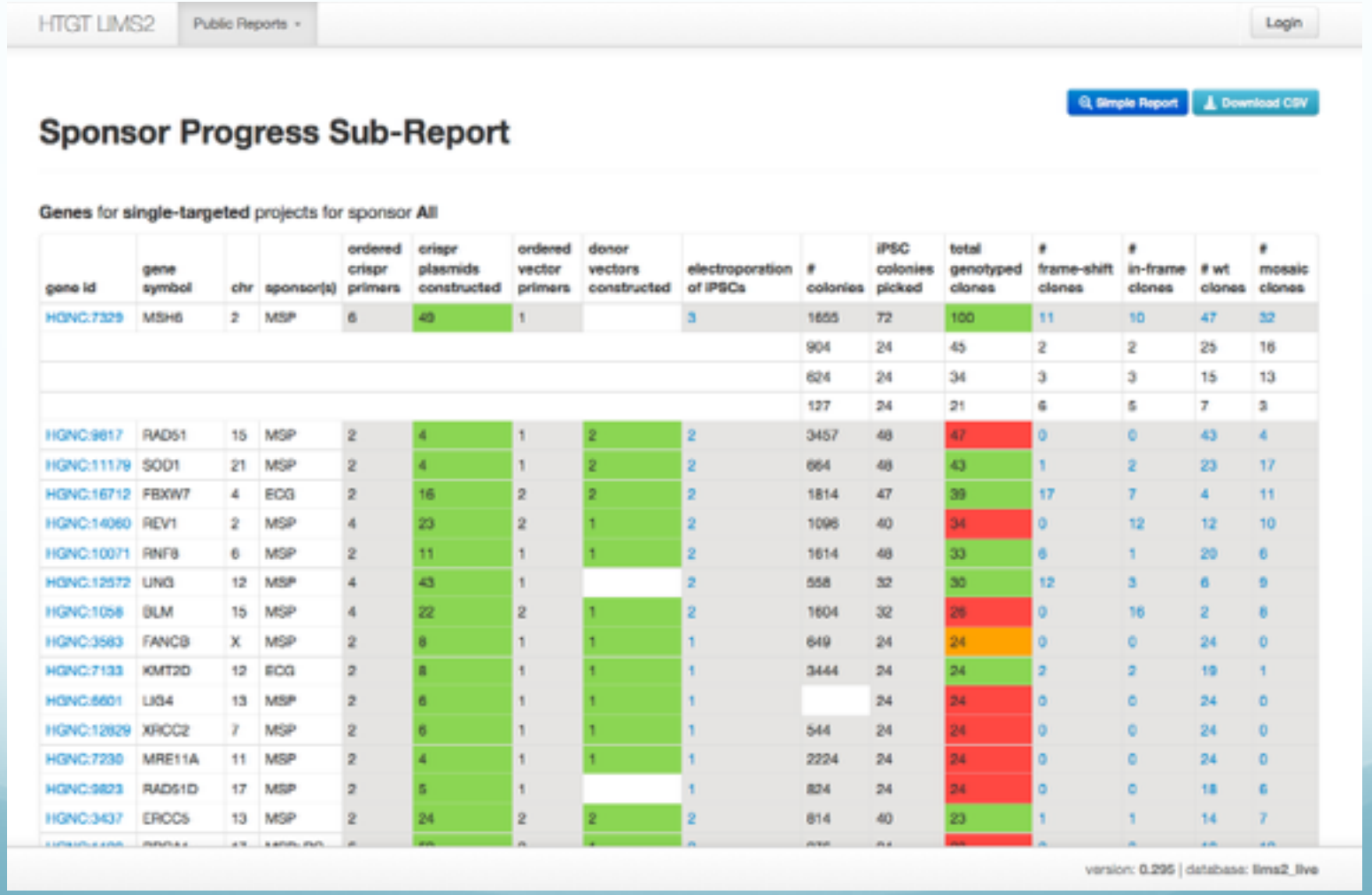
# LIMS2 Public Access



# Drill down ...


HTGT LIMS2
Public Reports
Login
Simple Report
Download CSV
Sponsor Progress Sub-Report
Genes for single-targeted projects for sponsor All
gene id | gene symbol | chr | sponsor(s) | ordered crispr primers | crispr plasmids constructed | ordered vector primers | donor vectors constructed | electroporation of iPSCs | # colonies | iPSC colonies picked | total genotyped clones | # frame-shift clones | # in-frame clones | # wt clones | # mosaic clones
HGNC:7329 | MSH6 | 2 | MSP | 6 | 49 | 1 | | 3 | 1655 | 72 | 100 | 11 | 10 | 47 | 32
HGNC:9817 | RAD51 | 15 | MSP | 2 | 4 | 1 | 2 | 2 | 3457 | 48 | 47 | 0 | 0 | 43 | 4
HGNC:11179 | SOD1 | 21 | MSP | 2 | 4 | 1 | 2 | 2 | 664 | 48 | 43 | 1 | 2 | 23 | 17
HGNC:16712 | FBXW7 | 4 | ECG | 2 | 16 | 2 | 2 | 2 | 1814 | 47 | 39 | 17 | 7 | 4 | 11
HGNC:14060 | REV1 | 2 | MSP | 4 | 23 | 2 | 1 | 2 | 1096 | 40 | 34 | 0 | 12 | 12 | 10
HGNC:10071 | RNF8 | 6 | MSP | 2 | 11 | 1 | 1 | 2 | 1614 | 48 | 33 | 6 | 1 | 20 | 6
HGNC:12572 | UNG | 12 | MSP | 4 | 43 | 1 | | 2 | 558 | 32 | 30 | 12 | 3 | 6 | 9
HGNC:1058 | BLM | 15 | MSP | 4 | 22 | 2 | 1 | 2 | 1604 | 32 | 26 | 0 | 16 | 2 | 8
HGNC:3583 | FANCB | X | MSP | 2 | 8 | 1 | 1 | 1 | 649 | 24 | 24 | 0 | 0 | 24 | 0
HGNC:7133 | KMT2D | 12 | ECG | 2 | 8 | 1 | 1 | 1 | 3444 | 24 | 24 | 2 | 2 | 19 | 1
HGNC:6601 | LIG4 | 13 | MSP | 2 | 6 | 1 | 1 | 1 | | 24 | 24 | 0 | 0 | 24 | 0
HGNC:12829 | XRCC2 | 7 | MSP | 2 | 6 | 1 | 1 | 1 | 544 | 24 | 24 | 0 | 0 | 24 | 0
HGNC:7230 | MRE11A | 11 | MSP | 2 | 4 | 1 | 1 | 1 | 2224 | 24 | 24 | 0 | 0 | 24 | 0
HGNC:9823 | RAD51D | 17 | MSP | 2 | 5 | 1 | | 1 | 824 | 24 | 24 | 0 | 0 | 18 | 6
HGNC:3437 | ERCC5 | 13 | MSP | 2 | 24 | 2 | 2 | 2 | 814 | 40 | 23 | 1 | 1 | 14 | 7
HGNC:1100 | BRCA1 | 17 | MSP; PG | 6 | 59 | 2 | 1 | 2 | 976 | 24 | 23 | 0 | 3 | 10 | 10
HGNC:9806 | RAD1 | 5 | MSP | 2 | 14 | 1 | 1 | 1 | 404 | 24 | 23 | 0 | 0 | 19 | 4
HGNC:886 | ATRX | X | MSP; SCE | 2 | 13 | 1 | 1 | 1 | 184 | 24 | 23 | 0 | 0 | 23 | 0
version: 0.295 | database: lims2_live

# … and fine detail



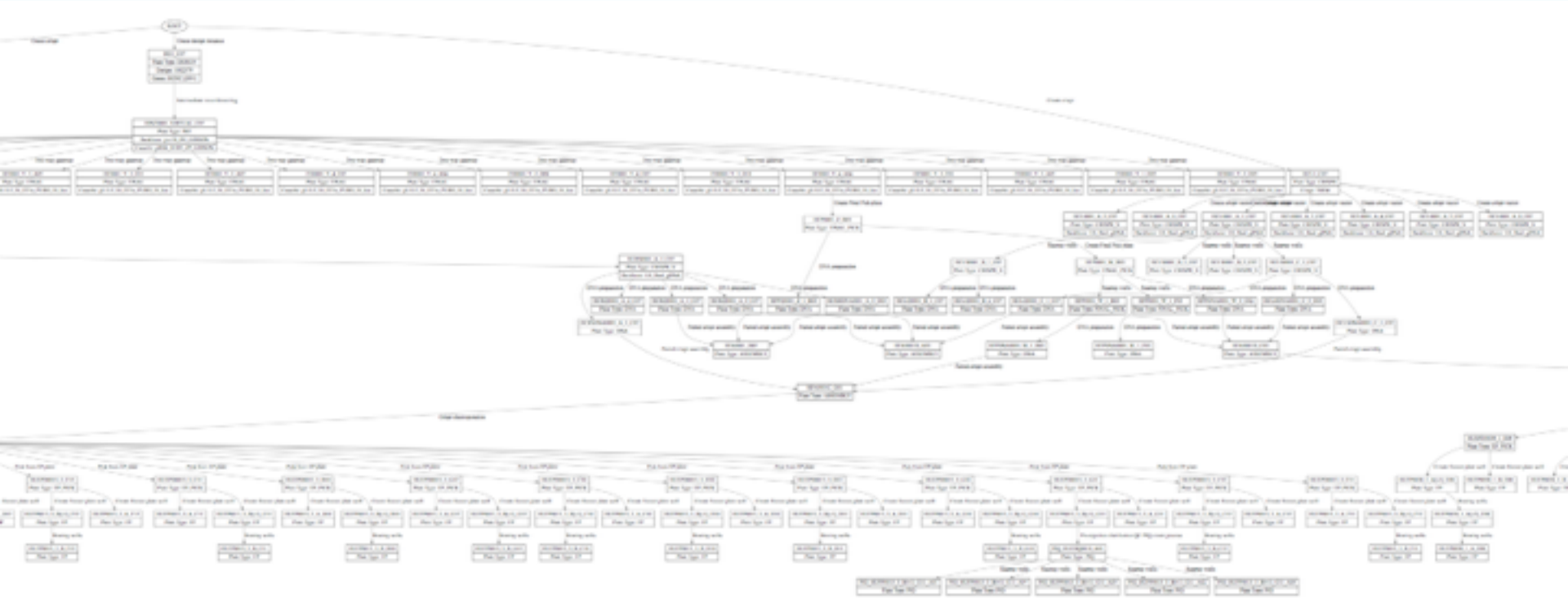
# CRISPR QC Trace Viewer

# How much data?

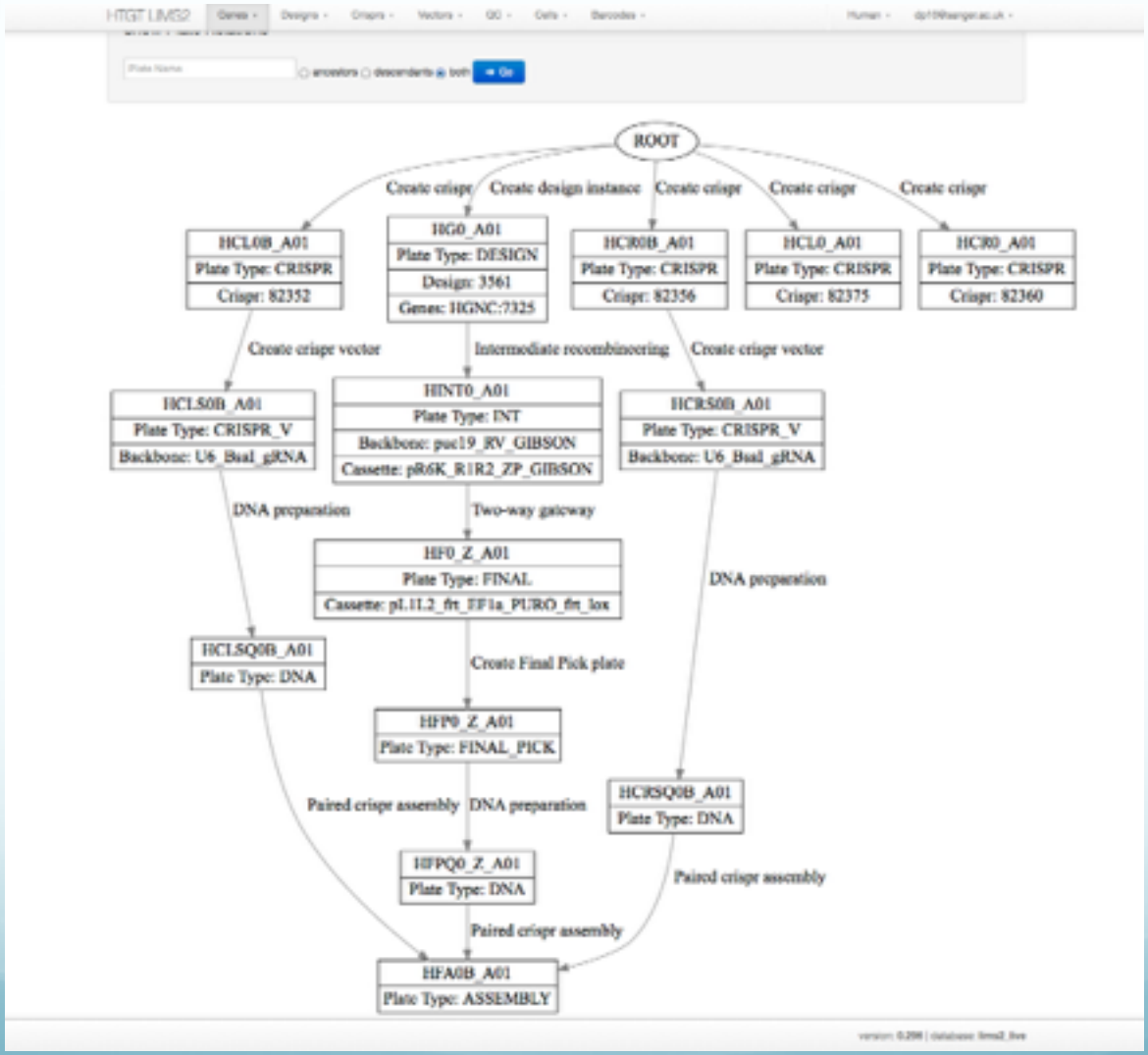
- LIMS2 is a well-relationship system
- As of today:
- 412,157 wells
- 20,000 design genes
  - Mouse 17,780*
  - Human 1,987

*Does not include data in HTGT2 that is not included in LIMS2

# How it looks to LIMS2



# A simpler example



# WGE

HTGT WGE | Home | CRISPR Finder | Gibson Designer | Help | About | Contact | Login with Google

**WTSI Genome Editing (WGE) is a website that provides tools to aid with genome editing of human and mouse genomes**

## CRISPR Finder

The CRISPR Finder will show CRISPR sites (paired or single) in and around genes. You can ask the finder to score the pairs for potential off-target sites, and browse individual and paired CRISPR sites using the Genoverse genome browser tool. We also provide the ability to find CRISPRs in genomic sequence or by gRNA:

**Find CRISPRs in our genome browser:**

**Find CRISPRs by gene using our table:**

**Find CRISPRs by 20bp gRNA:**

**Find CRISPRs in genomic sequence:**

**Find off-targets by sequence:**

## Gibson Designer

The Gibson Designer will find the oligos in either Human or Mouse genomes that can be used to create targeting vectors by Gibson assembly. The Gibson Designer matches the vector design with CRISPR sites appropriate for the creation of exon deletions.

If you use this site in your research, please cite:

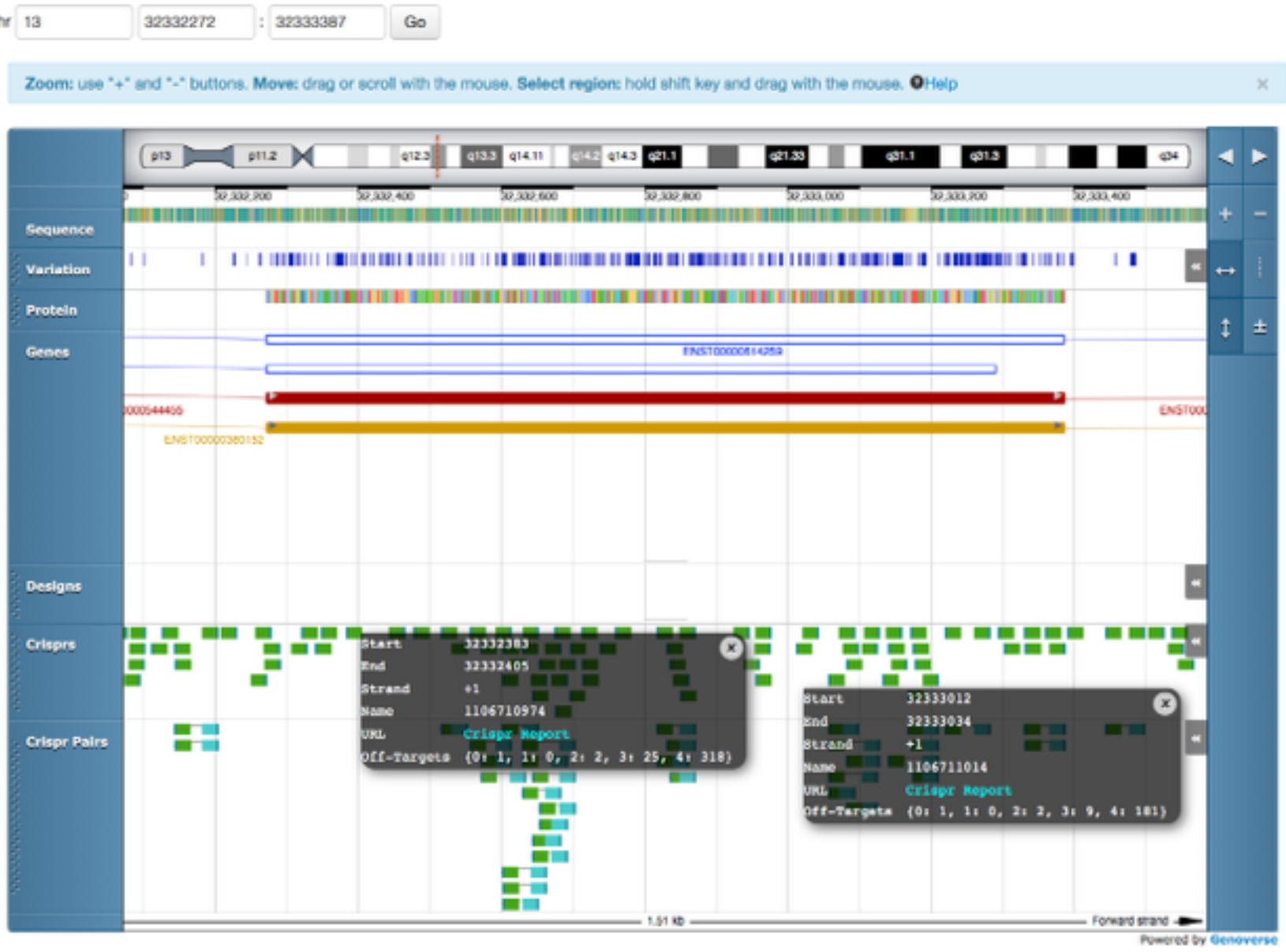
Bin Shen, Wensheng Zhang, Jun Zhang, Jiankui Zhou, Jianying Wang, Li Chen, Lu Wang, Alex Hodgkins, Vivek Iyer, Xingxu Huang & William C Skarnes (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. doi:10.1038/nmeth.2857

wellcome trust sanger institute

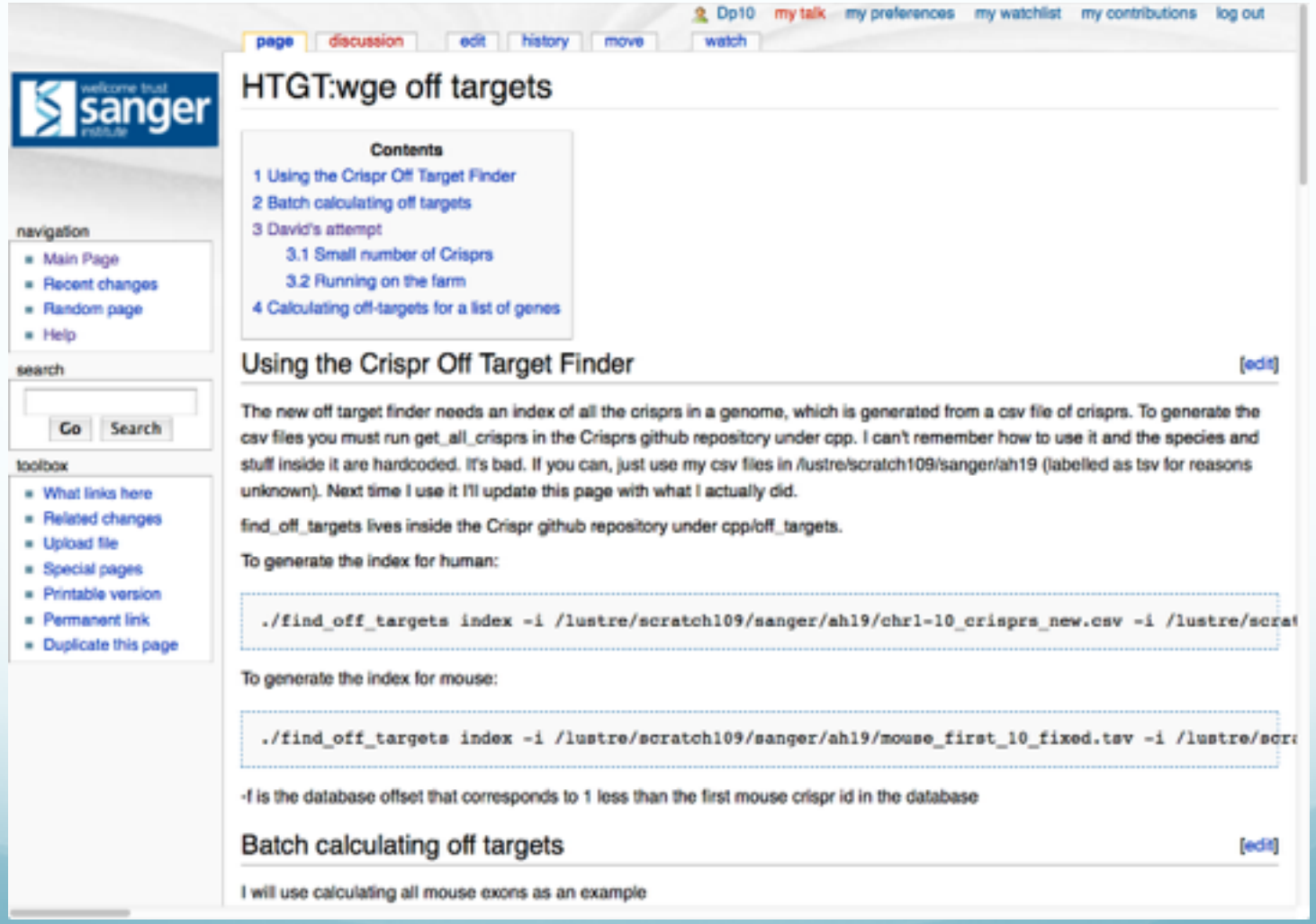
# WGE - Implementation

- Originally a Dancer app
- Now a Perl/Catalyst/DBIx::Class/PostgreSQL app
- Genome browser interface to a database of CRISPR sites in the Human and Mouse genomes (novel!)
- Uses an in memory index served by separate server
- Off-targets are calculated and stored
  - Mouse and Human exomes (+ 200 bp flank)
  - batch mode or via genome browser interface

# WGE Genoverse



# Wiki Documentation



Dp10 my talk my preferences my watchlist my contributions log out

page discussion edit history move watch

sanger institute

navigation
- Main Page
- Recent changes
- Random page
- Help

search
Go Search

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Duplicate this page

## HTGT:wge off targets

Contents
1 Using the Crispr Off Target Finder
2 Batch calculating off targets
3 David's attempt
3.1 Small number of Crisprs
3.2 Running on the farm
4 Calculating off-targets for a list of genes

### Using the Crispr Off Target Finder [edit]

The new off target finder needs an index of all the crisprs in a genome, which is generated from a csv file of crisprs. To generate the csv files you must run get_all_crisprs in the Crisprs github repository under cpp. I can't remember how to use it and the species and stuff inside it are hardcoded. It's bad. If you can, just use my csv files in /lustre/scratch109/sanger/ah19 (labelled as tsv for reasons unknown). Next time I use it I'll update this page with what I actually did.

find_off_targets lives inside the Crispr github repository under cpp/off_targets.

To generate the index for human:

```
./find_off_targets index -i /lustre/scratch109/sanger/ah19/chr1-10_crisprs_new.csv -i /lustre/scrat
```

To generate the index for mouse:

```
./find_off_targets index -i /lustre/scratch109/sanger/ah19/mouse_first_10_fixed.tsv -i /lustre/scr
```

-f is the database offset that corresponds to 1 less than the first mouse crispr id in the database

### Batch calculating off targets [edit]

I will use calculating all mouse exons as an example

# Future Plans

- Prepare users for the move to read-only HTGT
- Migration of LIMS2 to Ubuntu 12.04 or (preferably) 14.04
- Review of Deployment strategy for LIMS2 and WGE
  - Currently bespoke
  - Consideration of publicly available tools
    - Puppet, Chef etc.
- Extend functions of LIMS2 as a tracking system driven by user requirements

# People

- Informatics Group 2015
  - Anna Farne
  - Tiago Grego
  - David Parry-Smith
  - Saj Pereira
- Ex-members
  - Andrew Sparkes
  - Alex Hodgkins
  - Richard Easty
  - Vivek Iyer
- New role – Imits lead
  - Peter Matthews
- Scientists
  - Wendy Bushell
  - Barry Rosen
  - Mark Thomas
  - Bill Skarnes