Experimental tools: a new way to categorise transgenic alleles in FlyBase

Gillian Millburn and the FlyBase Consortium

Abstract

The long history of fly research plus the sophisticated range of applicable genetic engineering techniques mean that a large number of increasingly complex transgenic fly lines have been generated and described in the literature. While this rich genetic tool-kit helps to make *Drosophila melanogaster* an ideal model organism to answer a wide range of biological questions, it also creates a potential problem - how to find the most appropriate fly line for a particular experiment from the large set that are available. To help address this issue, FlyBase has recently introduced the 'experimental tool' data class. Reports have been generated for commonly used sequences with useful properties that are exploited to study the biological function of another gene product or a biological process. These include tools that enable a gene product to be detected (e.g. the **FLAG** tag, **EGFP**, **mCherry**), target a gene product somewhere specific within a cell (e.g. *nuclear localization signal*, *signal sequence*), drive expression in a binary system (e.g. **UAS**, **GAL4**) or enable clonal/conditional expression (e.g. **FLP**, **FRT**). Controlled vocabulary terms are used to describe the common uses for each tool (e.g. *epitope tag*, *green fluorescent protein*, *recombinase*). Linking the appropriate experimental tools to transgenic constructs and insertions allows researchers to more easily identify constructs and fly stocks that have the particular characteristics they are interested in. Experimental tools have been linked to all relevant transgenic constructs and alleles in the database and this curation is now included as a standard part of the curation of new genetic reagents.

Curation model

1. Make a new 'Experimental Tool' report for each 'flavour' of tool.

2. Describe what the tool is **used for** using a set of ontology terms.

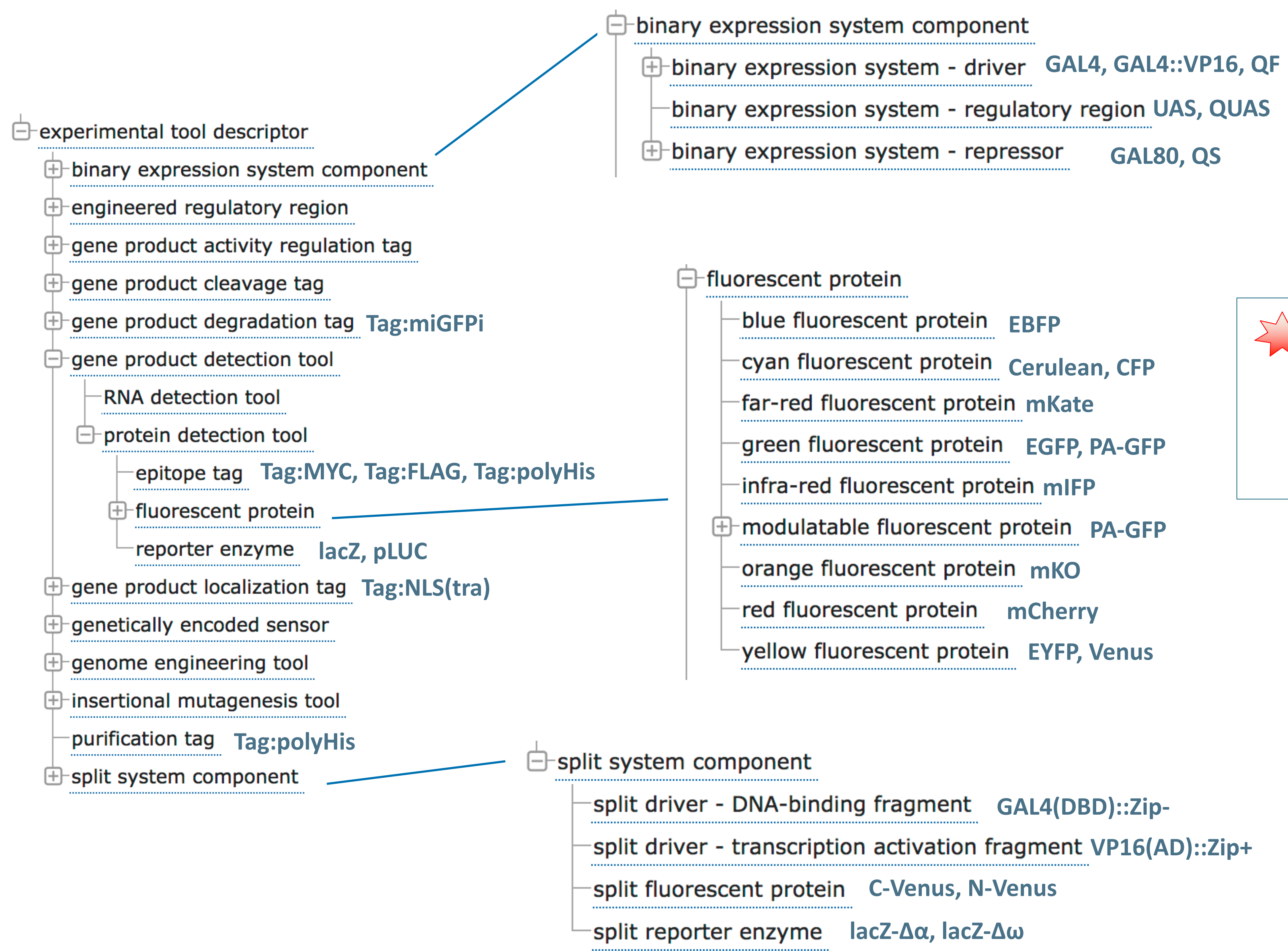| Experimental Tool	'Uses' ontology term(s)
EGFP	green fluorescent protein
mCherry	red fluorescent protein
lacZ	reporter enzyme
Tag:MYC	epitope tag
GAL4	binary expression system – driver
UAS	binary expression system - regulatory region
ArcLight	fluorescent protein / voltage sensor

3. Link tool(s) to the appropriate transgenes and alleles in the database.

Tools can be linked to each individual transgene/allele in four different ways, to build up a picture of what it is made up of.

Regulatory region	The tool corresponds to an engineered regulatory region which is used to drive expression of the gene product encoded by the transgenic construct/modified endogenous locus.
Encoded tool	The entire gene product encoded by the transgenic construct/modified endogenous locus acts as a tool.
Tagged with	The tool is fused to ('tags') another gene product whose biological function is being studied, and confers a novel property on that gene product.
Also carries	The tool does not form part of the gene product, but is 'carried in' the transgenic construct/modified endogenous locus.

'Uses' ontology

Main branches of **Uses** ontology and **examples** of **Tools** annotated with these terms

- experimental tool descriptor
  - binary expression system component
    - binary expression system - driver — **GAL4, GAL4::VP16, QF**
    - binary expression system - regulatory region **UAS, QUAS**
    - binary expression system - repressor **GAL80, QS**
  - engineered regulatory region
  - gene product activity regulation tag
  - gene product cleavage tag
  - gene product degradation tag **Tag:miGFPi**
  - gene product detection tool
    - RNA detection tool
    - protein detection tool
      - epitope tag **Tag:MYC, Tag:FLAG, Tag:polyHis**
      - fluorescent protein
        - blue fluorescent protein **EBFP**
        - cyan fluorescent protein **Cerulean, CFP**
        - far-red fluorescent protein **mKate**
        - green fluorescent protein **EGFP, PA-GFP**
        - infra-red fluorescent protein **mIFP**
        - modulatable fluorescent protein **PA-GFP**
        - orange fluorescent protein **mKO**
        - red fluorescent protein **mCherry**
        - yellow fluorescent protein **EYFP, Venus**
      - reporter enzyme **lacZ, pLUC**
  - gene product localization tag **Tag:NLS(tra)**
  - genetically encoded sensor
  - genome engineering tool
  - insertional mutagenesis tool
  - purification tag **Tag:polyHis**
  - split system component
    - split driver - DNA-binding fragment **GAL4(DBD)::Zip-**
    - split driver - transcription activation fragment **VP16(AD)::Zip+**
    - split fluorescent protein **C-Venus, N-Venus**
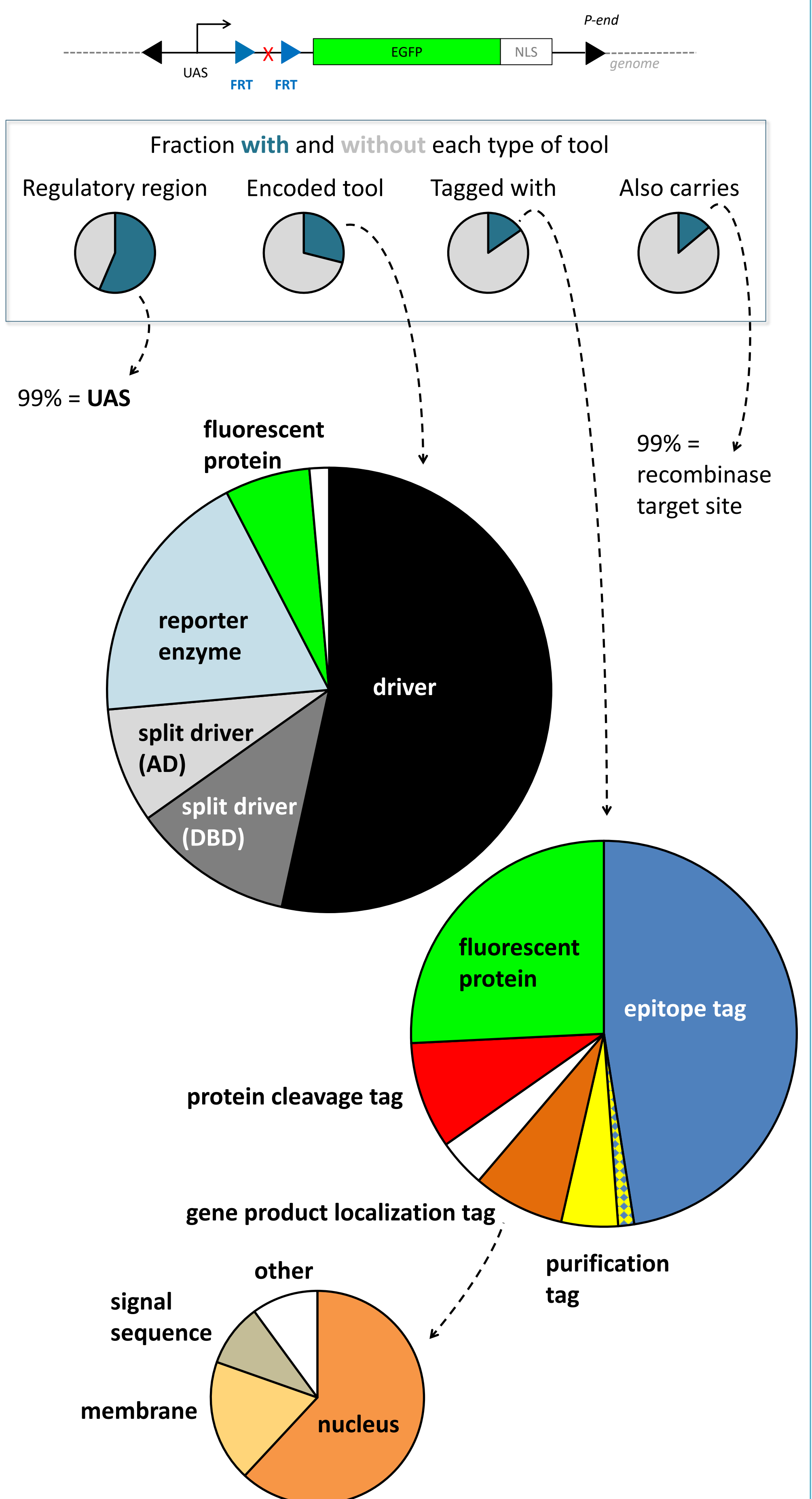   - split reporter enzyme **lacZ-Δα, lacZ-Δω**

Coming soon!
- cell ablation tool
- neuron activation tool
- neuron inhibition tool

The data

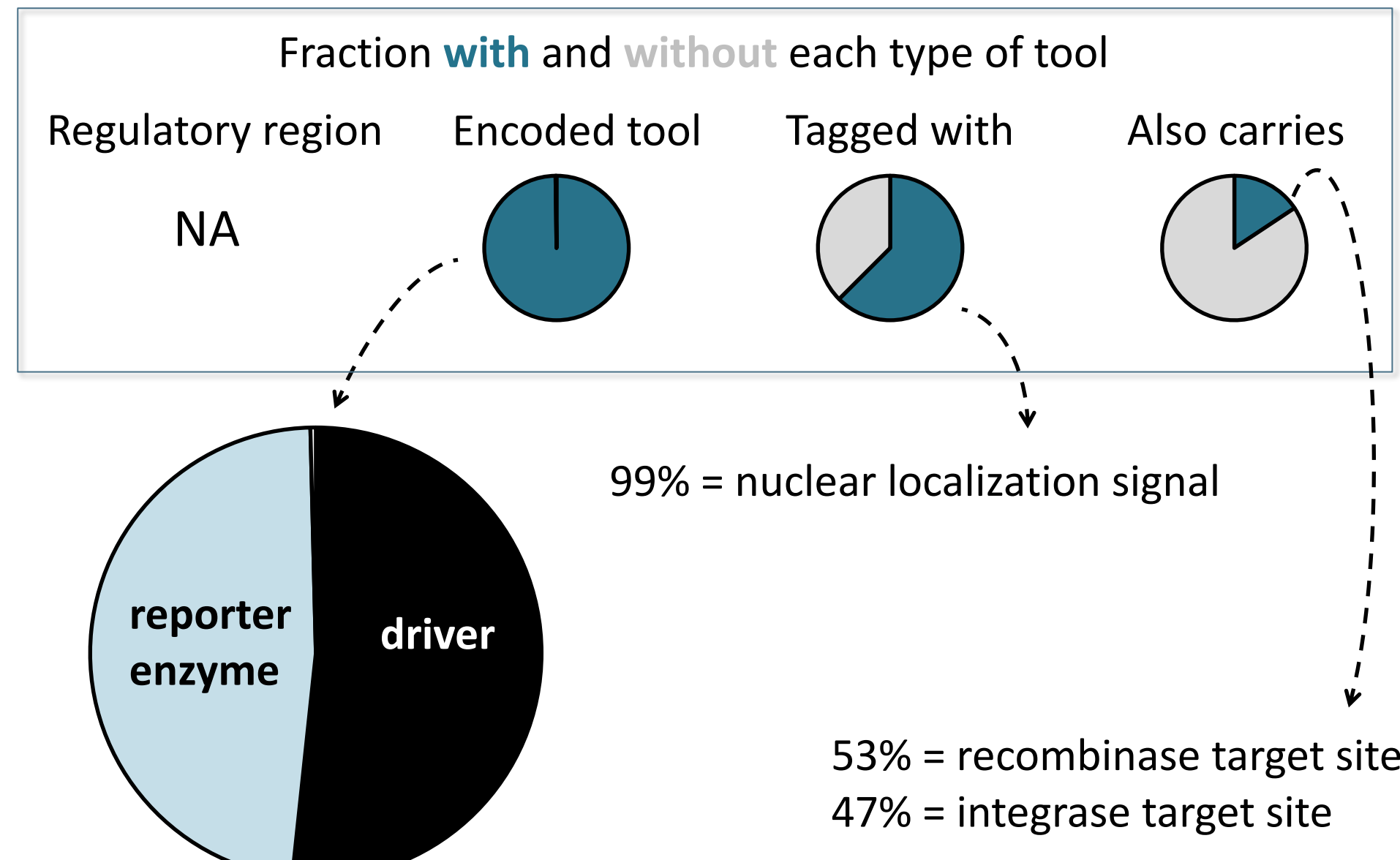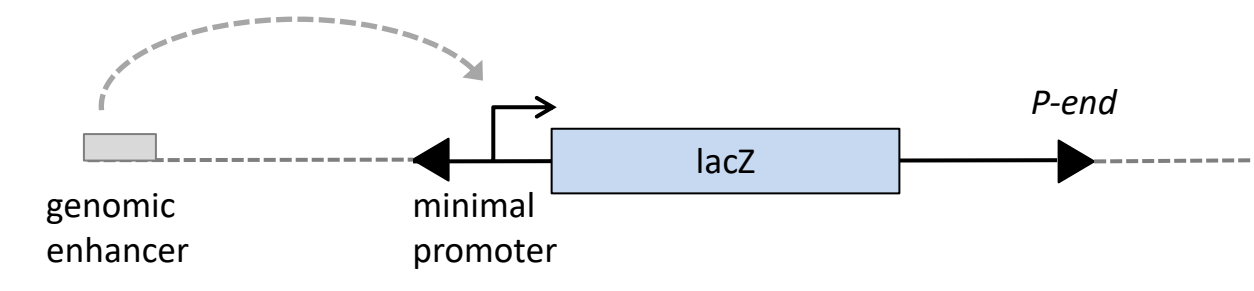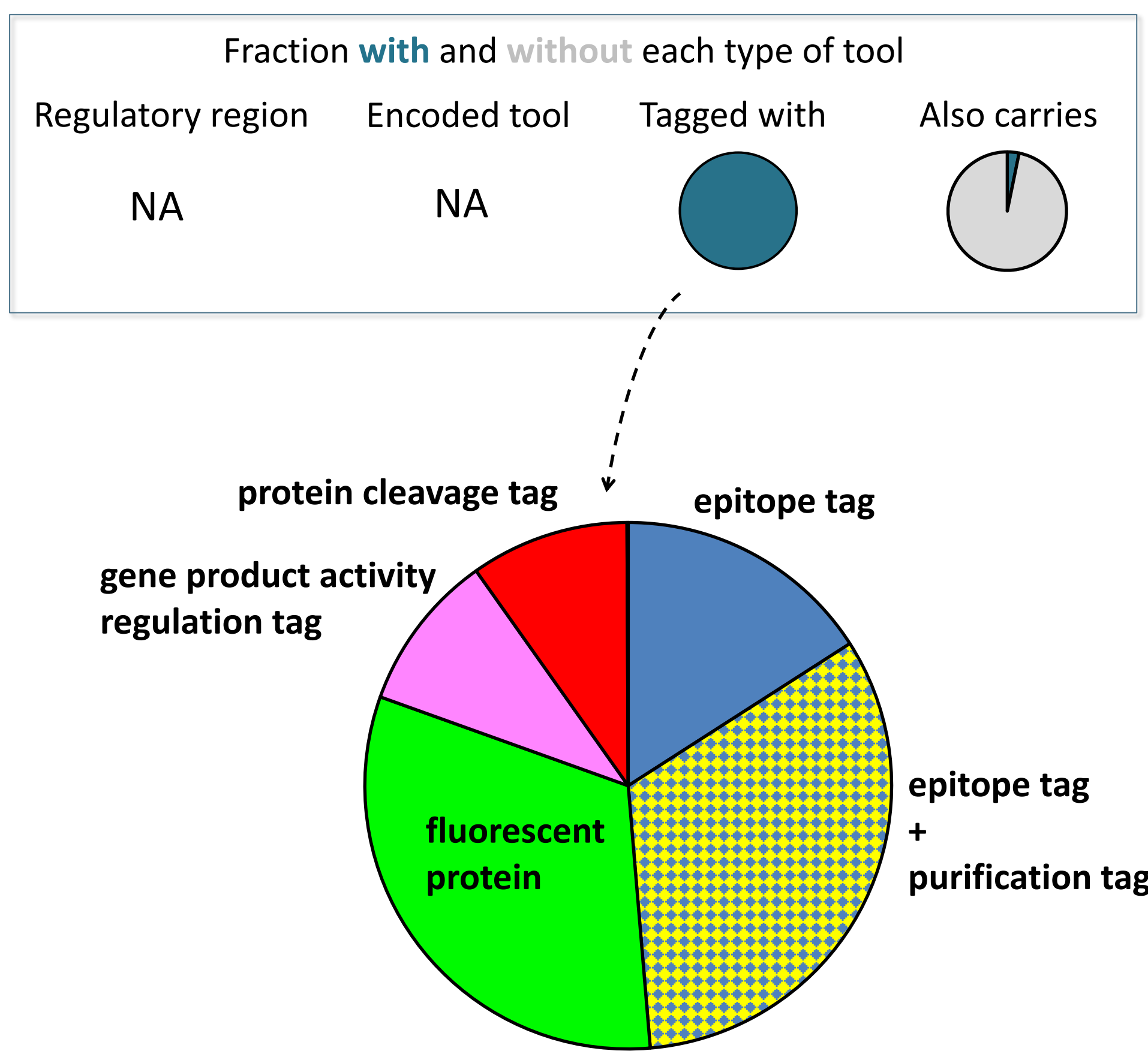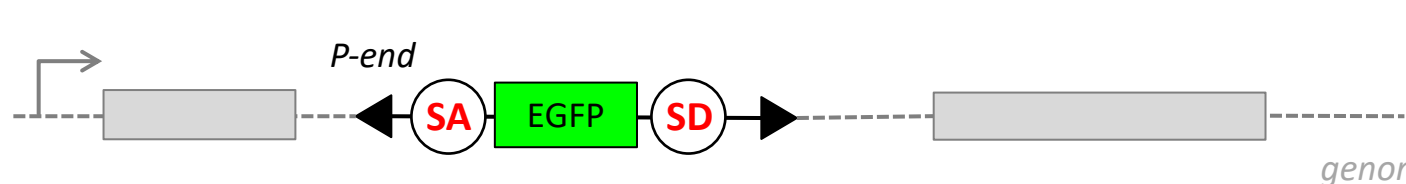### 'Promoter fusion' style transgenes — 126784

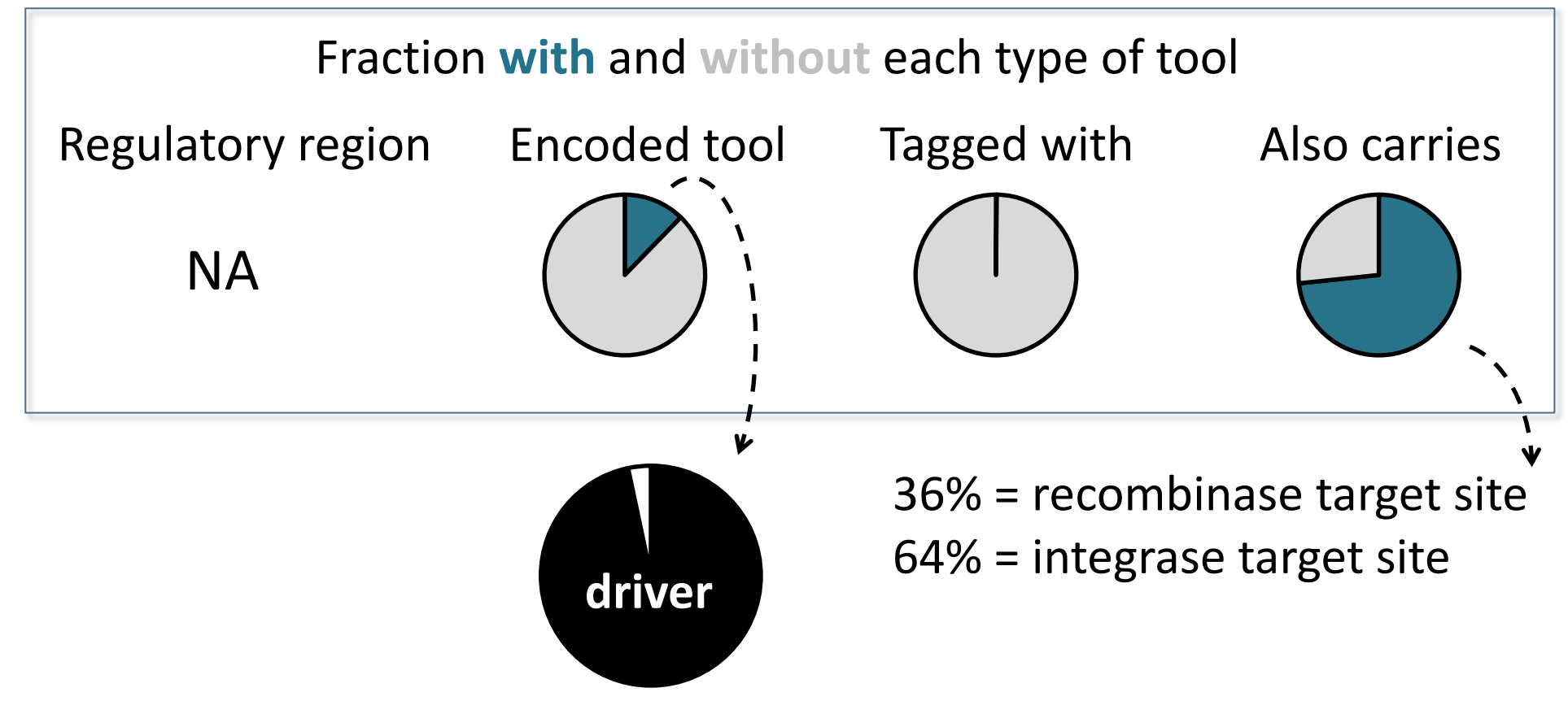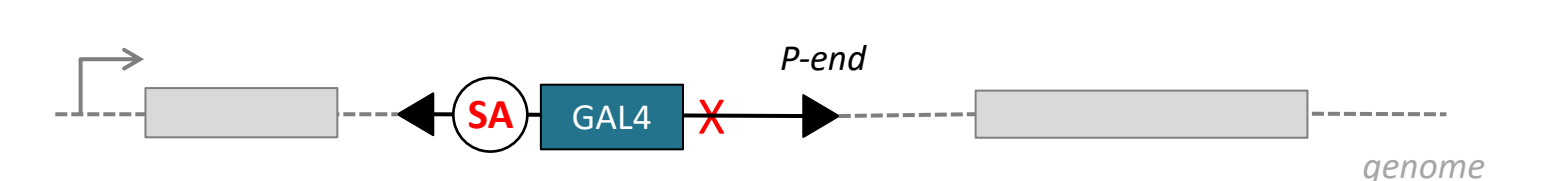- regulatory region that drives expression is included in the transgene

Fraction **with** and without each type of tool
- Regulatory region — 99% = **UAS**
- Encoded tool
- Tagged with — 99% = recombinase target site
- Also carries

Enhancer traps — 21293

Fraction **with** and without each type of tool
- Regulatory region — NA
- Encoded tool
- Tagged with — 99% = nuclear localization signal
- Also carries — 53% = recombinase target site / 47% = integrase target site

Gene traps — 15693

Fraction **with** and without each type of tool
- Regulatory region — NA
- Encoded tool — driver
- Tagged with
- Also carries — 36% = recombinase target site / 64% = integrase target site

Misexpression elements — 34697

Fraction **with** and without each type of tool
- Regulatory region — 99% = **UAS**
- Encoded tool — NA
- Tagged with — NA
- Also carries — 99% = recombinase target site

Protein traps — 2177

Fraction **with** and without each type of tool
- Regulatory region — NA
- Encoded tool — NA
- Tagged with
- Also carries
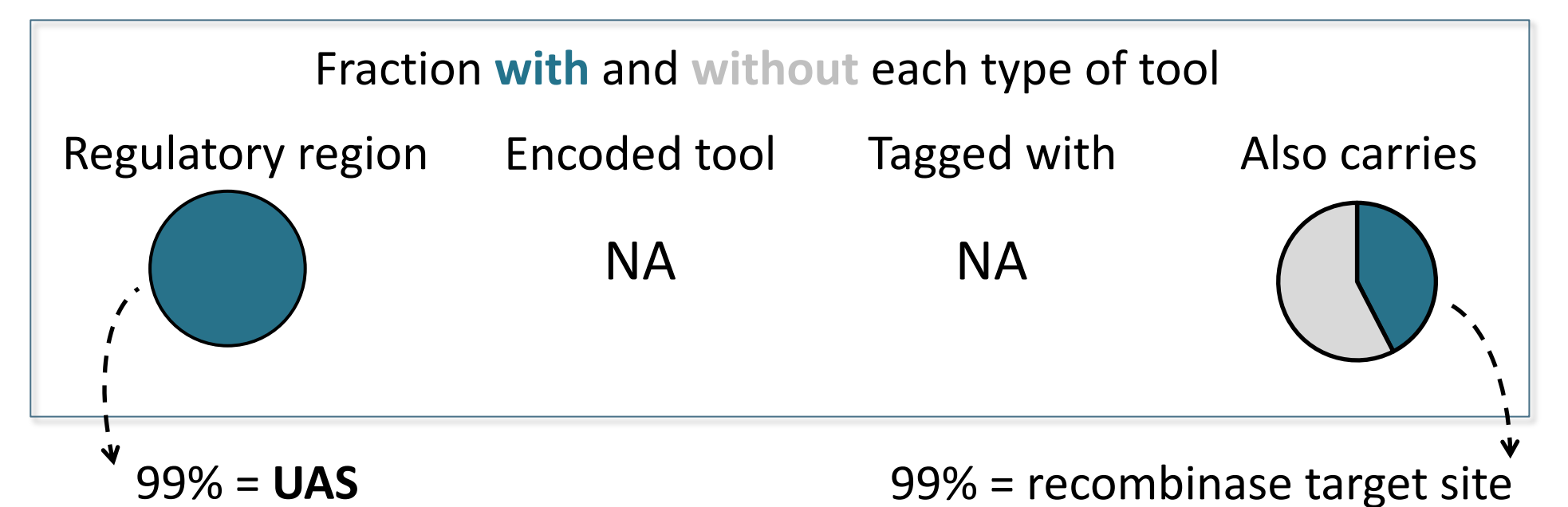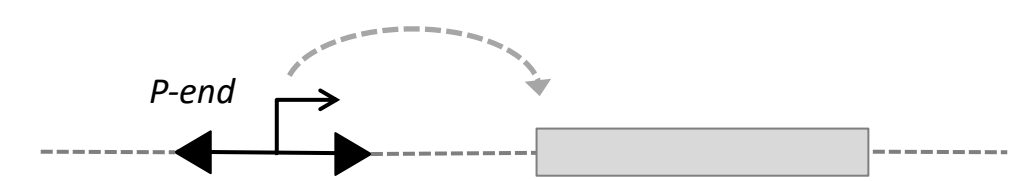
Genome engineering at endogenous loci — 2833

- mutation and/or insertion introduced direct into the genome using a technique such as CRISPR/Cas9 and/or homologous recombination

Fraction **with** and without each type of tool
- Regulatory region — single case = **UAS**
- Encoded tool — driver
- Tagged with
- Also carries — 61% = recombinase target site / 39% = integrase target site