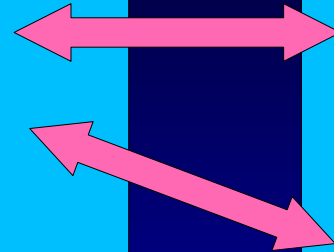


# Object relation in schema

Cell

Rearrangement

```
[
  {
    "cell_id": "AA001",
    "rearrangements": [
      "3141",
      "5926"
    ],
    "repertoire_id": "rep_1",
    "virtual_pairing": false
  },
  {
    "cell_id": "AA002",
    "rearrangements": [
      "5358",
      "9793"
    ],
    "repertoire_id": "rep_1",
    "virtual_pairing": false
  }
]
```



```
[
  {
    "sequence_id": "3141",
    "locus": "IGH",
    "sequence": "ACG...",
    "repertoire_id": "rep_1",
    "cell_id": "AA001"
  },
  {
    "sequence_id": "5926",
    "locus": "IGK",
    "sequence": "GCA...",
    "repertoire_id": "rep_1",
    "cell_id": "AA001"
  },
  {
    "sequence_id": "5358",
    "locus": "IGH",
    "sequence": "TCA...",
    "repertoire_id": "rep_1",
    "cell_id": "AA002"
  },
  {
    "sequence_id": "9793",
    "locus": "IGL",
    "sequence": "ACA...",
    "repertoire_id": "rep_1",
    "cell_id": "AA002"
  }
]
```

# Expression data in schema

## Cell

```
[
  {
    "cell_id": "AA001",
    "rearrangements": [
      "3141",
      "5926"
    ],
    "repertoire_id": "rep_1",
    "expression_study_method": "scTrans",
    "expression_tabular": [
      {
        "expression_marker": "PAX5",
        "expression_value": 6.0
      },
      {
        "expression_marker": "ACTB",
        "expression_value": 51.0
      }
    ],
    "virtual_pairing": false
  },
  ...
]
```

## Rearrangement

```
[
  {
    "sequence_id": "3141",
    "locus": "IGH",
    "sequence": "ACG...",
    "repertoire_id": "rep_1",
    "cell_id": "AA001"
  },
  {
    "sequence_id": "5926",
    "locus": "IGK",
    "sequence": "GCA...",
    "repertoire_id": "rep_1",
    "cell_id": "AA001"
  },
  ...
]
```

# Data provided by API

/v1/cell

```
[
  {
    "cell_id": "AA001",
    "rearrangements": [
      "3141",
      "5926"
    ],
    "repertoire_id": "rep_1",
    "expression_study_method": "scTrans",
    "expression_tabular": [
      {
        "expression_marker": "PAX5",
        "expression_value": 6.0
      },
      {
        "expression_marker": "ACTB",
        "expression_value": 51.0
      }
    ],
    "virtual_pairing": false
  },
  ...
]
```

/v1/rearrangement

sequence_id	locus	sequence	repertoire_id	cell_id
3141	IGH	ACG...	rep_1	AA001
5926	IGK	GCA...	rep_1	AA001
...				

(alternatively as JSON, see previous slide)

# On-disk option 1: Complete cell.json

cell.json

```
[  
  {  
    "cell_id": "AA001",  
    "rearrangements": [  
      "3141",  
      "5926"  
    ],  
    "repertoire_id": "rep_1",  
    "expression_study_method": "scTrans",  
    "expression_tabular": [  
      {  
        "expression_marker": "PAX5",  
        "expression_value": 6.0  
      },  
      {  
        "expression_marker": "ACTB",  
        "expression_value": 51.0  
      }  
    ],  
    "virtual_pairing": false  
  },  
  ...  
]
```

Bi-directional references are redundant and need to be kept consistent, thus only a uni-directional reference is provided.

rearrangement.tsv

sequence_id	locus	sequence	repertoire_id	cell_id
3141	IGH	ACG...	rep_1	AA001
5926	IGK	GCA...	rep_1	AA001
...				

\* abbreviated for readability, actual string would be "single-cell transcriptome"

# On-disk option 2: Separate expression table

cell.json

```
[  
  {  
    "cell_id": "AA001",  
    "repertoire_id": "rep_1",  
    "virtual_pairing": false  
  },  
  ...  
]
```

rearrangement.tsv

```
sequence_id locus sequence repertoire_id cell_id  
3141 IGH ACG... rep_1 AA001  
5926 IGK GCA... rep_1 AA001  
...
```

expression\_tabular.tsv

```
cell_id repertoire_id expression_study_method expression_marker expression_value  
AA001 rep_1 scTrans* PAX5 6.0  
AA001 rep_1 scTrans* ACTB 51.0  
...
```

\* abbreviated for readability, actual string would be "single-cell transcriptome"

# On-disk option 3: No cell.json

cell.json

```
[  
  {  
    "cell_id": "AA001",  
    "repertoire": "rep_1"  
  },  
  ...  
]
```

rearrangement.tsv

```
sequence_id locus sequence repertoire_id cell_id virtual_pairing  
3141 IGH ACG... rep_1 AA001 false  
5926 IGK GCA... rep_1 AA001 false  
...
```

expression\_tabular.tsv

```
cell_id repertoire_id expression_study_method expression_marker expression_value  
AA001 rep_1 scTrans* PAX5 6.0  
AA001 rep_1 scTrans* ACTB 51.0  
...
```

\* abbreviated for readability, actual string would be "single-cell transcriptome"

# On-disk option 4: Fancy structured format

cell.json

```
[  
  {  
    "cell_id": "AA001",  
    "repertoire": "rep_1"  
  },  
  ...  
]
```

rearrangement.tsv

```
sequence_id locus sequence repertoire_id cell_id virtual_pairing  
3141 IGH ACG... rep_1 AA001 false  
5926 IGK GCA... rep_1 AA001 false  
...
```

expression\_tabular.fancy-structured-format

```
repertoire_id: rep_1  
expression_study_method: scTrans*  
table(cell_id, expression_marker, expression_value): {  
  AA001 AA002 AA003 ...  
PAX5 6.0 ...  
ACTB 51.0 ...  
...  
}
```

\* abbreviated for readability, actual string would be "single-cell transcriptome"