Developing a Multiscale Neural Connectivity Knowledgebase of the Autonomic Nervous System

Fahim T. Imam [1,*], Thomas H. Gillespie [1], Ilias Ziogas [2], Monique C. Surles-Zeigler [1], Susan Tappan [3], Burak I. Ozyurt [1], Jyl Boline [4], Bernard de Bono [5], Jeffrey S. Grethe [1] and Maryann E. Martone [1]

[1] *FAIR Data Informatics Lab, Department of Neurosciences, School of Medicine, University of California San Diego, La Jolla, CA, USA*
[2] *SPARC K-Core, La Jolla, CA, USA*
[3] *Rock Maple Science, LLC, Hinesburg, VT, USA*
[4] *Informed Minds Inc, Walnut Creek, CA, USA*
[5] *Whitby et al., LLC, Indianapolis, IN, USA*

Correspondence*:
Fahim T. Imam
fimam@ucsd.edu

ABSTRACT

The Stimulating Peripheral Activity to Relieve Conditions (SPARC) program is a U.S. National Institutes of Health (NIH) funded effort to enhance our understanding of the neural circuitry responsible for visceral control. SPARC's mission is to identify, extract, and compile our overall existing knowledge and understanding of the autonomic nervous system (ANS) connectivity between the central nervous system (CNS) and end organs. A major goal of SPARC is to use this knowledge to promote the development of the next generation of neuromodulation devices and bioelectronic medicine for nervous system diseases. As part of the SPARC program, we have been developing SCKAN, a dynamic knowledge base of ANS connectivity that contains information about the origins, terminations, and routing of ANS projections. The distillation of SPARC's connectivity knowledge into this knowledge base involves a rigorous curation process to capture connectivity information provided by experts, published literature, textbooks, and SPARC scientific data. SCKAN is used to automatically generate anatomical and functional connectivity maps on the SPARC portal.

In this article, we present the design and functionality of SCKAN, including the detailed knowledge engineering process developed to populate the resource with high quality and accurate data. We discuss the process from both the perspective of SCKAN's ontological representation as well as its practical applications in developing information systems. We share our techniques, strategies, tools and insights for developing a practical knowledgebase of ANS connectivity that supports continual enhancement.

Keywords: knowledge representation, ontologies, knowledge graph, neural connectivity, autonomic nervous system

1 INTRODUCTION

An NIH Common Fund initiative, the Stimulating Peripheral Activity to Relieve Conditions (SPARC; `https://commonfund.nih.gov/sparc)`, is a large-scale program aiming to accelerate the development of therapeutic devices that modulate electrical activity in nerves to improve organ function. SPARC aims to foster the creation of robust scientific and technological foundations, paving the way for the development of more effective and targeted bio-electronic neuromodulation therapies for a wide range of diseases and conditions. The overarching goal of the first phase of the SPARC program (2017-2022) was to deliver detailed, integrated multi-species functional and anatomical maps and models of the autonomic nervous system (ANS) and its innervation of end organs. The second phase of SPARC (2022-2025) will deliver detailed anatomical and functional maps of the human vagus, as well as open-source libraries for new neuromodulation devices.

The SPARC Data and Resource Center (DRC) develops the infrastructure and standards for storing, reusing, and integrating data and other publicly available resources across the consortium. The DRC supports the data sharing platform, standards the SPARC knowledge graph, connectivity maps, and simulation tools for the mammalian nervous system (Osanlouy et al., 2021; Surles-Zeigler et al., 2022). Launched in July 2019, the SPARC Portal (`https://sparc.science/`) is the main entry point to SPARC data, maps, tools and community. The portal supports multiple search options across data, models, and simulations including faceted search, keyword search, and map-based query.

A key component of the SPARC project, the SPARC Connectivity Knowledgebase of the Autonomic Nervous System (SCKAN), is a semantic store containing detailed, formal knowledge about ANS connectivity between the central nervous system and end organs. Connectivity knowledge in SCKAN is derived from various authoritative sources, including SPARC investigators, anatomical experts, SPARC data, textbooks, and published literature. SCKAN is continually being enhanced with new connectivity statements derived from published literature using an NLP-backed knowledge curation pipeline. From extracting information from various sources to distilling the knowledge for its formal representation, the derivation of each connectivity statement in SCKAN involves a rigorous curation process. SCKAN is developed using semantic web technologies to support automated reasoning and semantic search.

SCKAN serves as a powerful resource to discover, and query ANS connectivity over multiple scales, species, and organs. SCKAN is being used within SPARC to generate interactive connectivity maps available via the SPARC Portal (`https://sparc.science/apps/maps?id=7d8ad680`). Figure 1 provides an example of SPARC's anatomical connectivity map showing different connectivity pathways in rats sourced from SCKAN.

In this article, we provide an overview of SCKAN and present details about its organization, curation, and population. We then demonstrate how SCKAN can be used to answer questions about ANS connectivity, including detailed routing of axons within individual nerves.

2 MATERIALS AND METHODS

We first give an overview of how connectivity information is modeled within SCKAN and then go through the steps of identifying and encoding connections for the population of SCKAN. Finally, we provide details on how to access SCKAN.

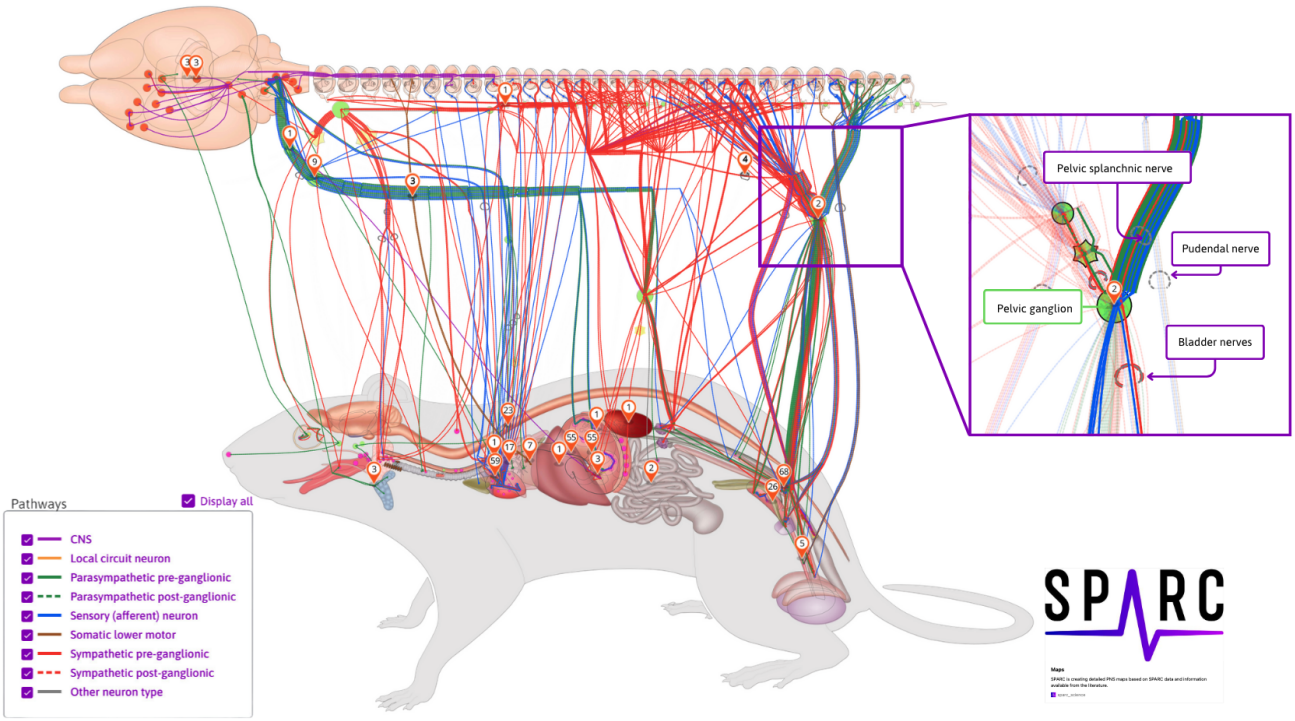

Figure 1. An example of SPARC's anatomical connectivity map demonstrating neuron connectivity sourced from SCKAN.

2.1 Modeling Connectivity

SCKAN models connectivity at the neuronal level, where connections are represented as locations of neuronal cell bodies, axon terminals, dendrites and axons (Figure 2). Each connection represents a population of PNS neurons with a given PNS phenotype (e.g. sympathetic pre-ganglionic) that projects to a given set of targets, where the exact cell type identity may not be known. By modeling connections as constituents of neuron populations, it becomes possible to assign phenotypes such as anatomical location to individual parts of a neuron. A connection from region A to region B traveling via nerve C is therefore represented as: neuron population (N) with somata in anatomical location A, axon terminals in structure B and axons traveling via nerve C. This relationship is further extended to support multiple locations [$A_{i\ldots n}$ to $B_{i\ldots n}$ via $C_{i\ldots n}$] in order to represent the distribution present. SCKAN further models individual axons as a set of segments which capture detailed routing via anatomical structures or nerves. We utilize the axon sensory terminal to indicate the peripheral process of pseudo-monopolar dorsal root ganglia connections.

Figure 3 provides examples of two neuron populations in SCKAN reflecting their projection paths. On the left, we have the pathway of a preganglionic neuron population originating at the superior salvatory nucleus and terminating at the pterygopalatine ganglion. The axons reach their target via the facial nerve, then the greater petrosal nerve, and the vidian nerve. A given population subserves a single PNS phenotype, e.g., sympathetic pre or post ganglionics, but may have multiple origins, primarily used to indicate that a connection arises from multiple spinal cord segments, and multiple terminal locations. On the right of Figure 3, we have an example of a sympathetic preganglionic population with multiple origins and terminals. SCKAN currently does not model whether a neuron population projecting to multiple regions does so via branching axons or whether distinct subpopulations project to distinct targets. The connectivity of a single neuron population includes distinct anatomical structures along the projection paths traversed by

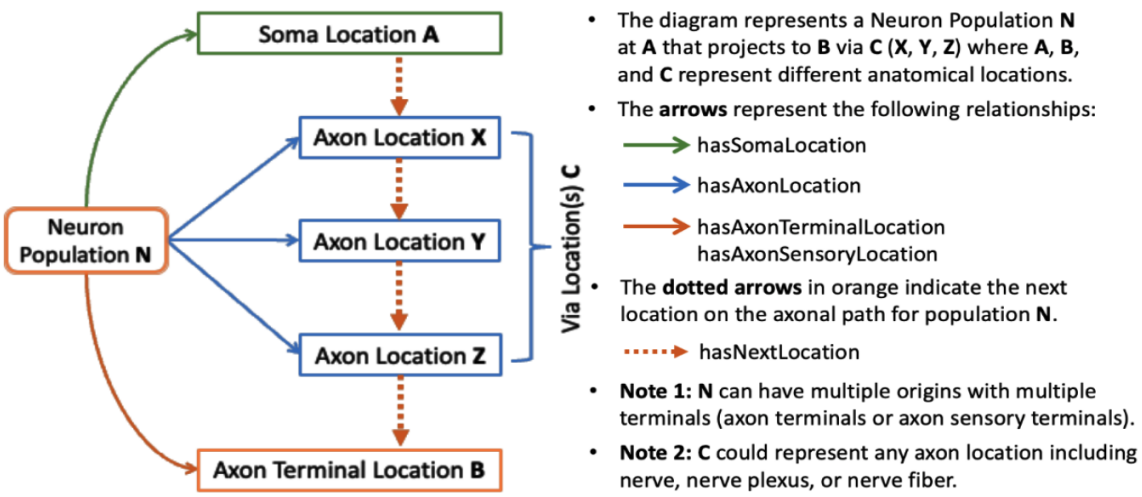

Figure 2. A connectivity model of a neuron population based on its axonal projection path.

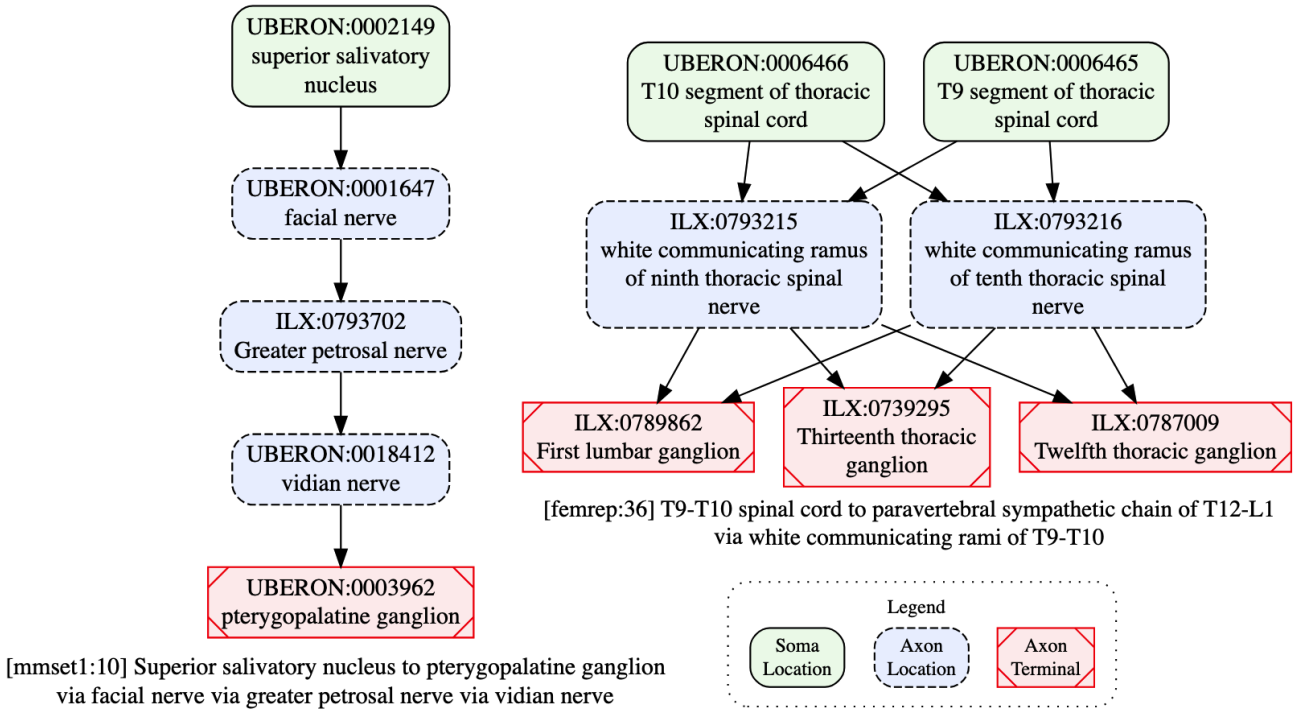

Figure 3. Examples of two neuron populations in SCKAN with projection paths from origin to destination.

a neuron population from its origin to its destination. In addition to specifying the anatomical locations of neuronal segments, each neuron population can be further described by specifying the following phenotypic properties, e.g.:

- ANS subdivision: sympathetic or parasympathetic, preganglinic or post-ganglionic
- Circuit role: intrinsic, motor, sensory, or projection
- Functional circuit role: whether the connection is excitatory or inhibitory
- Species and sex: the specificity of species and sex the connection was observed in

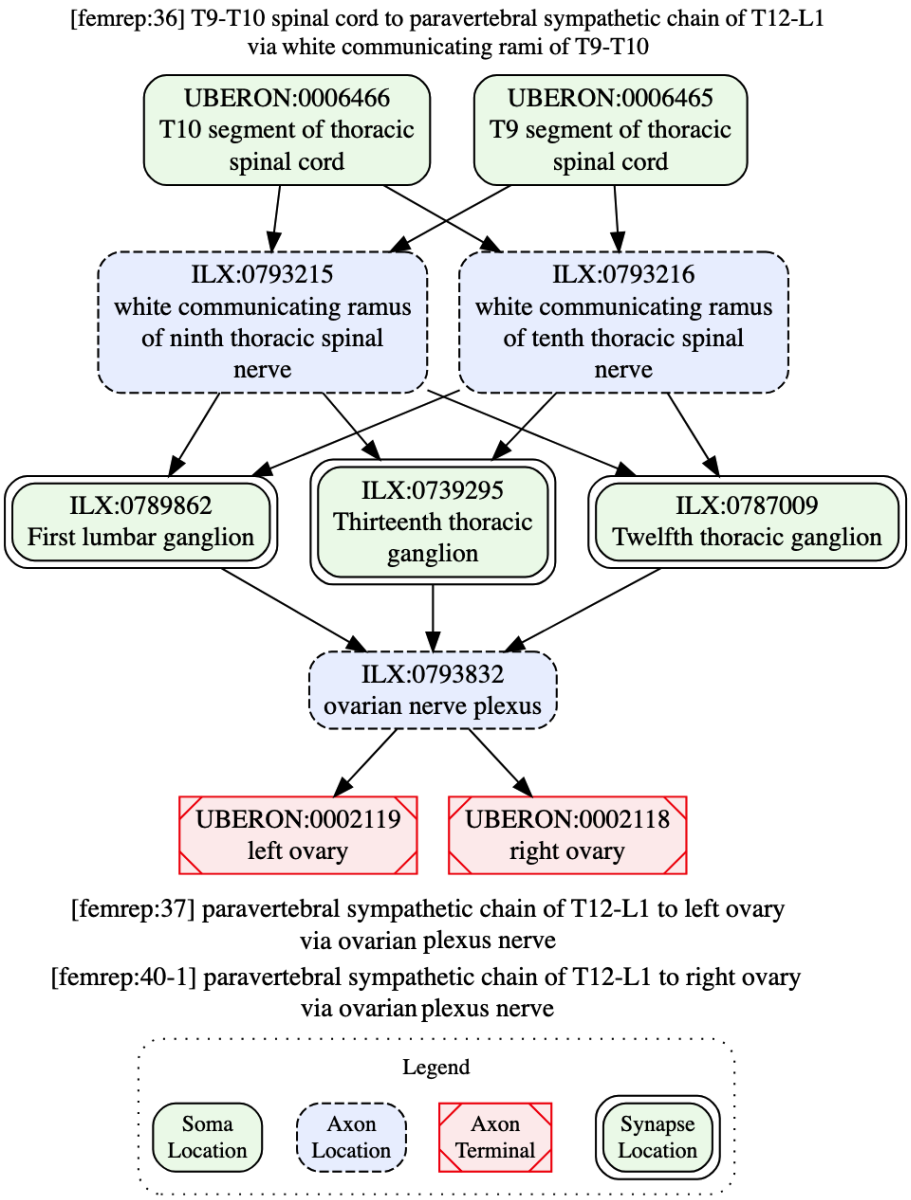

Figure 4. An example of a connectivity circuit formed by the synaptic connections between pre- and post-ganglionic populations in SCKAN.

SCKAN facilitates the modeling of connectivity circuits by specifying synaptic forward connections between first-order and second-order neuron populations. The preganglionic neuron populations, originating in the CNS and terminating in ANS ganglia, represent the first-order connections in SCKAN. The postganglionic populations, originating in the ANS ganglia and terminate in some part of a target organ, represent second-order connections. When a forward connection is specified between two populations, the synapse location within the circuit can be identified as the common structure(s) where the preganglionic connection terminates (the location of the axon presynaptic element) and the postganglionic connection originates. Figure 4 illustrates an example of a connectivity circuit in SCKAN formed by a synaptic forward connection between a sympathetic preganglionic population (femrep:36) and two postganglionic populations (femrep:37 and femrep:40-1) observed in female rats.

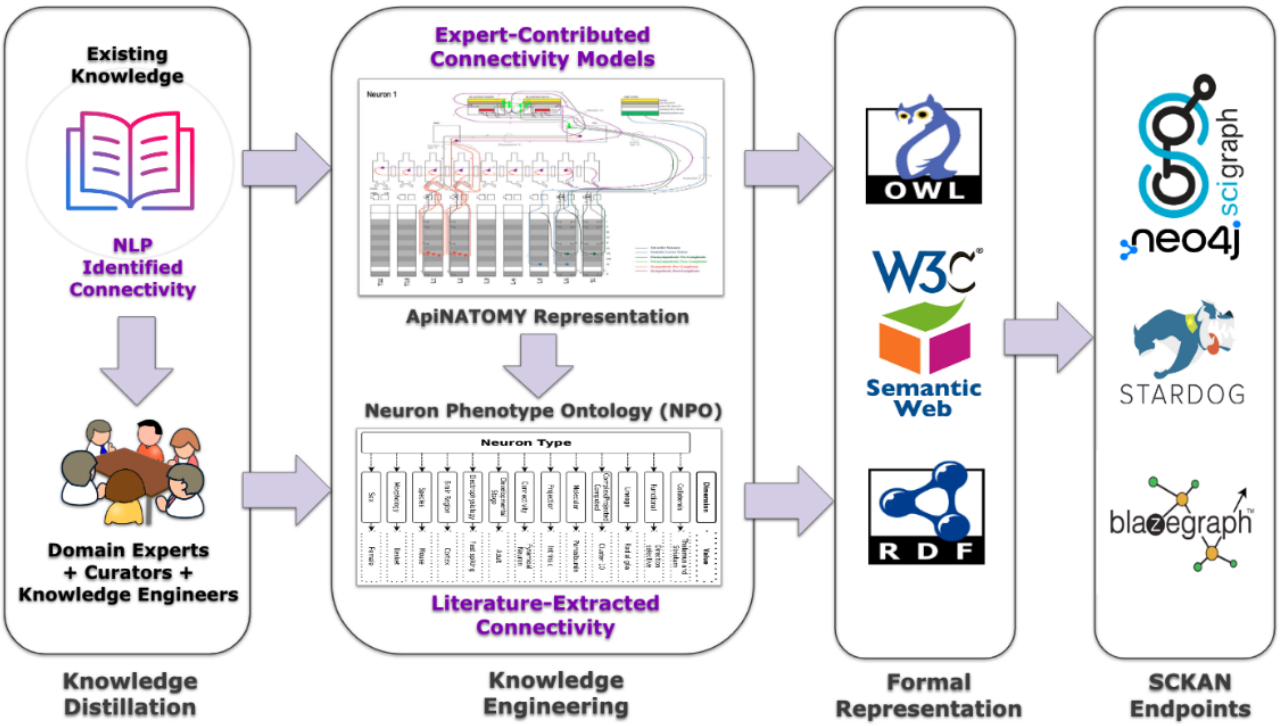

Figure 5. A high-level depiction of SCKAN's knowledge construction flow.

Extracting Connectivity Knowledge. Connectivity information contained in SCKAN is derived from two principle sources: (a) Expert-contributed connectivity models, and (b) Literature/textbook-extracted connectivity. Figure 5 provides a high-level depiction of SCKAN's knowledge construction flow, starting from the extraction of connectivity knowledge from these sources to the curation and transformation of this knowledge into formal representations, and finally making the knowledge base available via different graph database endpoints. We discuss the steps involved in the knowledge construction flow in the following sections.

2.2 Knowledge Distillation

The initial population of SCKAN was largely driven by detailed circuitry models of ANS connectivity associated with specific organs in Table 1. Each model was commissioned by the SPARC project based on relevance to SPARC Phase I and was constructed through consultations with domain experts. Note that as the focus of SPARC is on peripheral connectivity of the ANS, these models generally include the amount of CNS connectivity necessary to represent interactions with ANS populations, but do not aim to be comprehensive for CNS circuitry in the spinal cord and brainstem (or beyond to higher-order structures).

To ensure that SCKAN and the resulting flat maps were comprehensive of major autonomic connections, we developed an NLP-based pipeline (Fig 1) to identify papers in the scientific literature containing connectivity information (Ozyurt et al., 2021; Menke et al., 2020). The connectivity relation extraction classifier was trained to recognize sentences containing two or more peripheral anatomical structures related through connectivity, e.g., "... stimulation of sensory afferents from the trigeminal nerve results in parasympathetic vasodilation of the cerebral vasculature via interactions with the facial nerve and SPG" (White et al., 2021). The initial ANS connectivity relation extraction was conducted across the entirety of the full length open access subset of Pub Med Central (PMC) articles as well as additional sources to which

Model Name	Description
Keast Model of Bladder Innervation	Neural circuits in the central and peripheral nervous system innervating the urinary bladder and urethra. Observed predominantly in rodents.
SAWG Model of Bronchomotor Control	Neural circuits for visceral afferents from, and para/sympathetic effectors to, the lower airways.
Bolser-Lewis Model of Defensive Breathing	Neural circuits controlling defensive breathing with components under the influence of the Superior Cervical Ganglion. Observed predominantly in rodents.
SAWG Model of the Descending Colon	Neural circuits innervating the descending colon. Includes sympathetic and visceral sensing of a prototypical segment of distal colonic muscle layers, blood supply, Peyer's patches and enteroendocrine cells. Observed predominantly in felines.
UCLA Model of the Heart	Neural circuits innervating the heart observed in dogs, domestic pigs, and humans. Includes the sympathetic, parasympathetic, and intrinsic sensing of the four cardiac chambers, aorta, and pulmonary trunk.
SAWG Model of the Pancreas	Neural circuits controlling the Pancreas.
SAWG Model of the Stomach	Neural circuits controlling the stomach.
SAWG Model of the Spleen Innervation	Neural circuits in the CNS and PNS innervating the spleen.

Table 1. Table of neural control circuits based on various models and contributors

we had full text access through publisher agreements. The connectivity relation extraction data sources include the following:

- Full-length open-access subset of PMC: 4 million papers (downloaded on November 2020)
- Pubmed abstracts: 31.8 million abstracts (2021 baseline)
- Elsevier articles from 221 journals: 290,734 papers

This initial run resulted in several thousand sentences that had to be reviewed (see next section) After the initial run, the pipelines are run monthly on new full length PMC open access articles. These new publications can either be articles published since the last run or full texts that became available of older articles. Generally, these monthly runs result in < 50 new sentences for review.

Connectivity sentences were imported into a custom tool, OSCAP (The Open SAWG Connectivity Authoring Pipeline). Sentences and the papers from which they were extracted were reviewed by curators to determine whether there was new connectivity information that needed to be added to SCKAN according to the decision tree shown in Figure 6. Curators checked the contents of SCKAN using a query interface, SCKAN Explorer, described in a later section. The curators were individuals with a neuroscience background, familiar with the contents of SCKAN and basic knowledge engineering. Curators did not necessarily have a deep knowledge of the ANS, but they were competent to interpret connectivity information contained within experimental papers. Curators first determined whether the information presented within the sentence was within scope, i.e., ANS mammalian connectivity. Any sentences that were determined to have been returned in error, i.e., they did not contain connectivity information, were flagged for feedback to further train the algorithm. Relevant sentences/papers were then checked against the

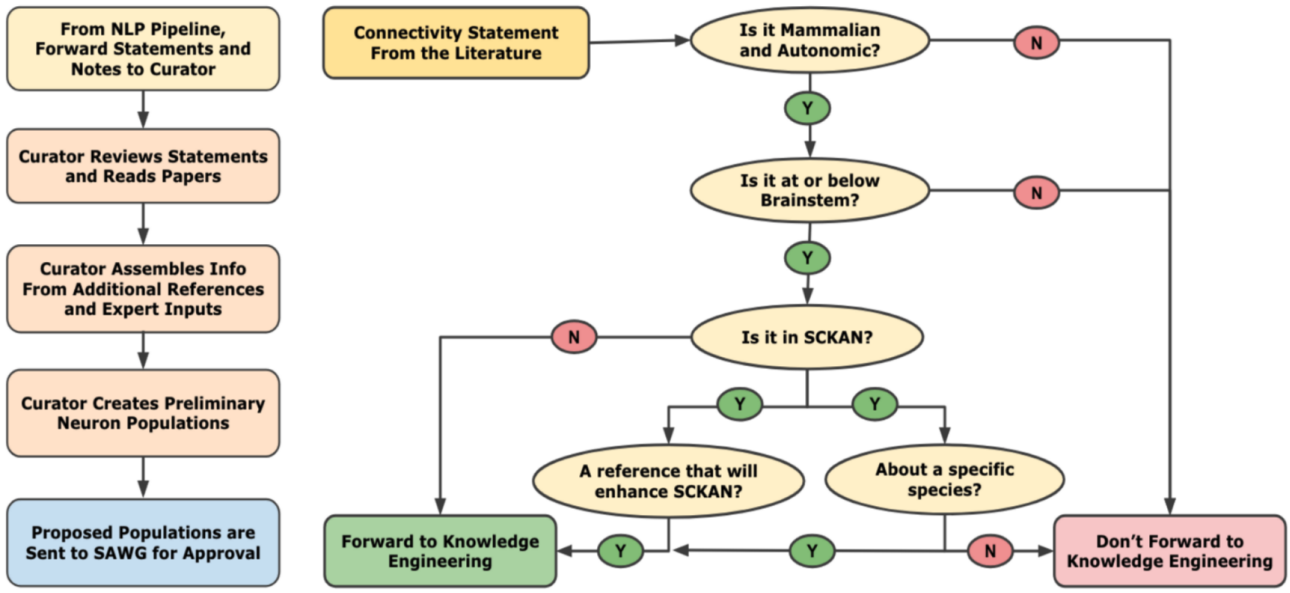

Figure 6. On the left: A decision tree for the knowledge curators to propagate connectivity knowledge from literature sources identified by the NLP pipeline. On the right: a typical workflow of a knowledge distillation process by SCKAN curators.

contents of SCKAN to see if that information was already present in the database. If the answer was yes, curators checked to see if new information about the connection was available, e.g., verifying an existing connection in another species. If this information was not in SCKAN, the sentence was flagged and the curator would add the connection, usually by tracking down additional references.

Once the curator identified a set of likely connections, a domain expert in the SPARC Anatomical Working Group (SAWG), a group of anatomical experts recruited by SPARC to provide anatomical expertise where needed, reviewed the evidence and provided input on whether these connections should be added to SCKAN. Experts provided input as to whether the proposed connections were well established or, for less well known connections, if the evidence provided was strong. If the evidence was less compelling, a decision was made by the curators and experts as to whether it should be included in the knowledge base. In these cases, any projections added were flagged with an alert explaining any concerns about the nature of the evidence.

We found during the course of the project that the most efficient way for domain experts to review proposed additions for SCKAN was to distill the proposed neuron population into a single sentence of the form "Structure A projects to Structure B via route C". This sentence was followed by evidence provided in the paper or papers on which the statement was based. In this way, the curator was explicit about the proposed connections and the expert could indicate whether they agreed with the statement (or parts of the statement) based on their expert knowledge and the evidence provided. The form of the sentence also directly mapped onto the underlying data model of SCKAN (see Section 2.1), facilitating communication with the knowledge engineer. This distillation process therefore provided the critical bridge between the domain expert and the knowledge engineer by providing the information in a way that could be easily reviewed and checked for accuracy by both.

Once the curator and domain experts agreed on the neuron populations to be added, the curator prepared them for handoff to the knowledge engineers. This process involved expressing each connection in terms

of anatomical locations of soma, axon segment and terminals, mapping anatomical locations to UBERON or Interlex identifiers (see section 2.4.2), and providing phenotype, species, and sex information. All connectivity statements were accompanied by provenance indicating their source. Acceptable sources include scientific articles, textbooks or expert knowledge. Each of these sources was mapped to the appropriate identifier: DOI or PMID, ISBN, or ORCIDs respectively. Curators could also add additional notes to any population to provide more information.

2.3 Knowledge Engineering

During the curation process, the curator consulted with the Knowledge Engineer about how to represent more challenging aspects of representation, e.g., branching, unidentified nerves. Over time, a set of rules were developed that were included in a curation manual to ensure that they were applied consistently. The curators were then tasked with performing initial knowledge engineering steps by mapping the entities within the distilled statements into formal knowledge statements. The majority of connectivity statements distilled by the knowledge curators and currently in SCKAN were entered using a standardized Google Sheets template, set up by the knowledge engineers ([Link to an example template]). This template provided a semi-formal structure for the curators to record connectivity statements according to standard, SCKAN-specific phenotypic dimensions, along with detailed provenance and curation notes for each neuron population. Curators utilize the commenting and tagging features of Google Sheets to consult with the SAWG and knowledge engineers whenever they encounter questions regarding any of their entries. Each connectivity statement recorded by the curators must be approved by SAWG before it can be added to SCKAN. Specifically, SAWG also provides feedback on the anatomical terms used in the statements. Once the curators have finalized their work, the curated Google Sheets are used to automatically transform the data into formal, machine-processable connectivity statements for SCKAN. Note that this process is now being replaced with the Composer-based workflow described below.

Over time, we found that the distillation and knowledge engineering processes were very error prone and very time consuming, e.g., all IDs had to be supplied manually. To streamline the distillation process, we developed the SCKAN Composer (RRID:SCR024441) which supports the entire pipeline, replacing both OSCAP and spreadsheets, and providing support for steps such as ontology ID look-ups. Composer was created to reduce the human effort required to identify, validate, and construct connectivity relationships and circuits, both initially and over time. The user interface of Composer allows the curator to draft a preliminary A to B via C statement, add tags, notes sections, provenance hyperlinks, species and sex identification. After initial distillation, the curator crafts the path trajectory utilizing anatomical entities retrieved from a data table that includes ontological IDs and synonyms. Composer automatically generates the stereotyped 'A-to-B-via-C' statement and a population diagram for review by the anatomical experts based on the entities selected by the curator. The production version of this tool was just released and is now supporting population of SCKAN (`https://composer.scicrunch.io/`. It will be further detailed in a subsequent publication.

2.4 Formal Representation

Once the pre-engineering steps were completed, the populations were handed off to the Knowledge Engineers for formal encoding. SCKAN is formally represented as an ontology using OWL-DL (`http://www.w3.org/TR/owl2-overview/`), a widely adopted formalism for ontology development. It leverages the Neuron Phenotype Ontology (NPO) as its foundational data model. NPO, developed as part of the BRAIN/BICCN and SPARC initiatives, provides a FAIR approach for representing and classifying neuron types based on their phenotypic properties (Gillespie et al., 2022). The notion of neuron types in

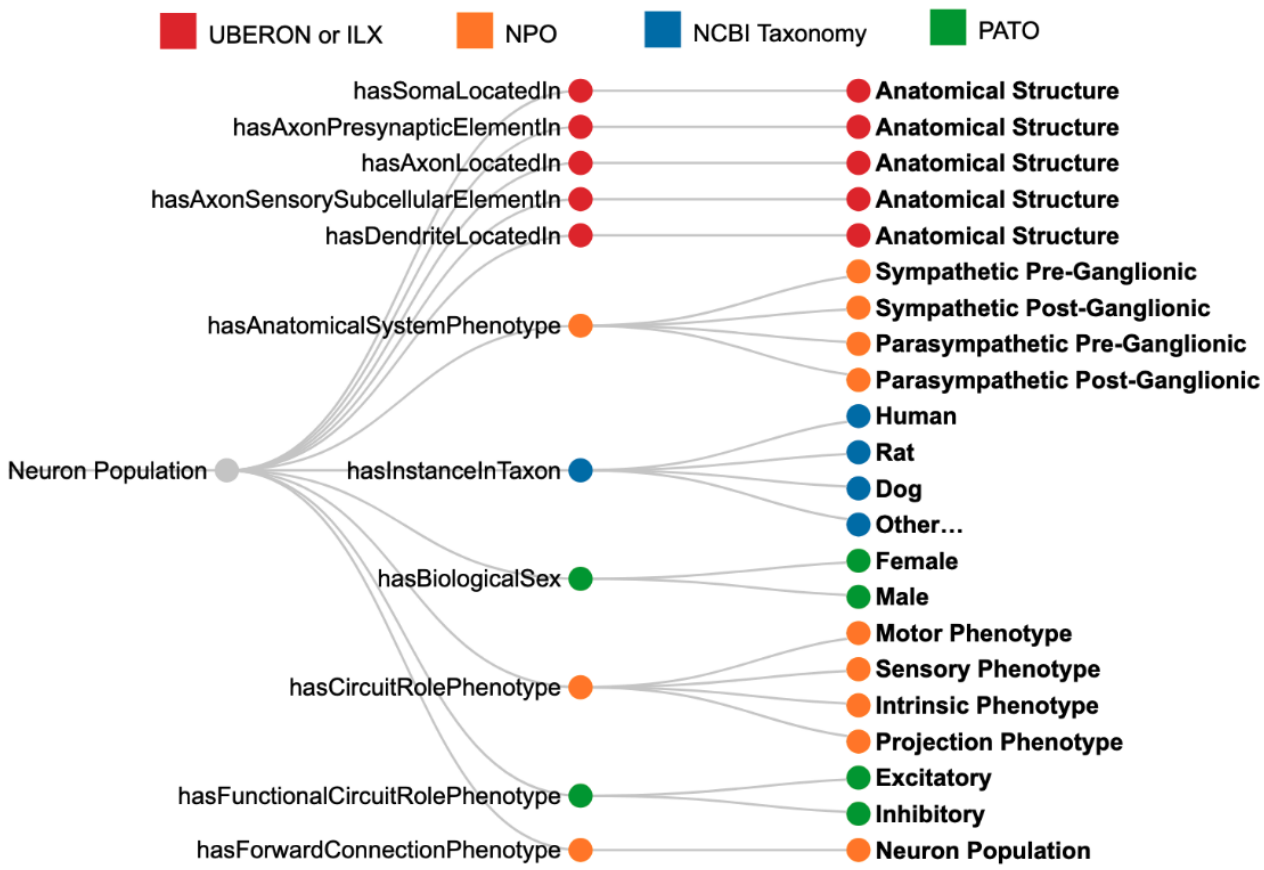

Figure 7. SCKAN-specific subset of properties from NPO and their possible values

NPO is generalized in SCKAN to represent region-to-region connectivity at the neuron population level. The OWL representation of a neuron population in SCKAN denotes a theoretical grouping of neurons based on shared phenotypic properties without specifying individual neuron types. SCKAN employs a subset of phenotypic properties from NPO to define its neuron populations as OWL classes. OWL enables SCKAN to be verified for consistency, infer implicit connectivity knowledge, and classify its populations through automated reasoning.

A class representing a neuron population in SCKAN must have its locational phenotypes associated with at least one origin and one destination. Figure 7 illustrates the SCKAN-specific subset of NPO's additional phenotypic properties and their possible values (property ranges). Adhering to FAIR principles, SCKAN derives phenotypic values from different community ontologies and vocabulary sources, as indicated by distinct node colors. Anatomical entities required to model locational phenotypes are primarily sourced from UBERON (refer to Section 2.4.2 for more details on vocabulary management).

Each neuron population class in SCKAN is assigned a unique URI identifier and is accompanied by a set of metadata annotations. These include human-readable labels from the curators, curatorial notes, and provenance information, with direct links to the source articles or artefacts from which curators extracted the connectivity information. Each neuron population class in SCKAN is represented as the equivalent to the intersections of phenotypic relations using OWL class expressions. For instance, the class "Prostate:24," which represents the neuron population "L2-L5 spinal cord to inferior mesenteric ganglion via lumbar

splanchnic nerve," is defined as equivalent to the OWL expression as shown in Listing 1 using Machester OWL syntax[**?**.

```
Class: prostate:24
Annotations:
  rdfs:label "neuron type prostate 24",
  skos:prefLabel "L2-L5 spinal cord to inferior mesenteric ganglion via lumbar splanchic nerve",
  ilxtr:literatureCitation <https://doi.org/10.1002/1097-0045(20000801)44:3%3C225::AID-PROS7%3E3.0.CO
      ;2-G>
EquivalentTo:
  ilxtr:NeuronSparcNlp and (hasInstanceInTaxon some 'Canis familiaris')
  and (hasBiologicalSex some male) and (not (hasBiologicalSex some female))
  and (hasAnatomicalSystemPhenotype some 'Sympathetic Pre-Ganglionic phenotype')
  and (hasSomaLocatedIn some ('part of' some 'L2 segment of lumbar spinal cord'))
  and (hasSomaLocatedIn some ('part of' some 'L3 segment of lumbar spinal cord'))
  and (hasSomaLocatedIn some ('part of' some 'L4 segment of lumbar spinal cord'))
  and (hasSomaLocatedIn some ('part of' some 'L5 segment of lumbar spinal cord'))
  and (hasAxonLocatedIn some ('part of' some 'lumbar splanchnic nerve'))
  and (hasAxonPresynapticElementIn some ('part of' some 'inferior mesenteric ganglion'))
  and (hasForwardConnectionPhenotype some 'neuron type prostate 25')
```

Listing 1. OWL Annotations and Equivalent Class Expression for a SCKAN neuron population

While the OWL axioms defining neuron populations in SCKAN provide a model for each neural segment, they cannot be used to determine the exact routing of the axonal course between the origin and destination, especially when multiple intermediate structures are involved. The OWL axioms alone cannot capture the complexity of such pathways.

```
prostate:24 a owl:Class ;
skos:prefLabel "L2-L5 spinal cord to inferior mesenteric ganglion via lumbar splanchic nerve" ;
ilxtr:neuronPartialOrder
    ( "blank"
        ( UBERON:0006447                # L5 segment of lumbar spinal cord
            ( UBERON:0018683            # lumbar splanchnic nerve
                ( UBERON:0005453 ) ) )  # inferior mesenteric ganglion
        ( UBERON:0006450                # L2 segment of lumbar spinal cord
            ( UBERON:0018683 ) )        # lumbar splanchnic nerve
        ( UBERON:0006449                # L3 segment of lumbar spinal cord
            ( UBERON:0018683 ) )        # lumbar splanchnic nerve
        ( UBERON:0006451                # L4 segment of lumbar spinal cord
            ( UBERON:0018683 ) )        # lumbar splanchnic nerve
    );
```

Listing 2. Specifying axonal course of locations using RDF list representation.

To represent the routing of the axonal segments comprising the entire axon segments and its branches, we use an RDF list representation to encode the hierarchical structure of routing among the anatomical entities traversed by the neuron populations. To facilitate this representation, we define a special property called 'ilxtr:neuronPartialOrder', which associates a neuron population with its partial order tree, representing the axonal course from the origin to the destination. This approach allows for the modeling of nested structures and hierarchical relationships that are essential for accurately depicting complex axonal pathways in SCKAN. The code segment in Listing 2 demonstrates a nested structure of representing the axonal pathways for a neuron population (prostate:24) in turtle format. The hierarchical structures represented in

the code can be queried using SPARQL to produce an adjacency list. The adjacency list is the primary way to render pathway diagrams.

The expert-contributed models in SCKAN were originally represented using ApiNATOMY (Kokash and de Bono, 2021), a sophisticated knowledge modeling framework to represent biological connectivity. ApINATOMY includes a suite of tools to produce anatomical schematics and connectivity diagrams of neural processes. ApiNATOMY provides a representation significantly more expressive than NPO for describing physical anatomy and the detailed routing of neural pathways located within larger tissue units (de Bono et al., 2022). In the context of neurons, this expressiveness allows for the attachment of additional information to individual axons and dendrites with a concomitant increase in complexity of the knowledge graph. Achieving similar granularity with NPO would require explicit modeling of axon and dendrite classes, a complexity we have deliberately avoided in order to simplify the queries. Apart from the partial ordering information for the neural pathways, the additional anatomical detail and adjacency knowledge provided by ApiNATOMY were not incorporated into SCKAN. Each expert-contributed model represented in ApiNATOMY was carefully transformed into an NPO-based representation using the SCKAN-specific subset of properties shown in Figure 7. SCKAN currently provides a unified representation of connectivity for both ApiNATOMY-based expert-contributed models and literature-extracted connections, enabling a uniform way to query connections from both sources. However, the detailed ApiNATOMY models remain accessible and can be queried via the SciGraph endpoint.

2.4.1 Modeling Species, Sex, and Laterality

Connectivity information from the literature pipeline outlined above has specific species and sex information associated with it when this information is available. SCKAN is designed according to an open world assumption, that is, lack of information does not imply that the information does not exist unless specified as such. Thus, for species information, assignment of a specific species is done through the ‘hasInstanceInSomeTaxon‘ relationship. The relationship connotes that a given connection was documented in a particular species but its presence or absence in another species is unknown. Neuron populations entered via the ApiNATOMY expert models often did not come with specific provenance for each connection. As SCKAN focuses exclusively on mammalian connectivity, these populations are coded as ‘hasInstanceInSomeTaxon some mammal’.

The ‘hasBiologicalSex’ relation is used to indicate sex differences in a given projection. If no sex is assigned, the projection is assumed to exist in both males and females. If a sex is explicitly assigned, it connotes that there is a sex difference between males and females, and we close the world by adding an explicit negation axiom for the other sex (See Line 12 in Listing 1 for an example).

Scientists are uncovering laterality differences between connections in certain species where there is an asymmetry in a given connection between the left and right sides of the body. Laterality differences are not encoded via relationships but via assignment of an explicit laterality label to the anatomical structures themselves. If a connection only exists on one side, e.g., in the sympathetic ovarian innervation via the superior ovarian nerve in the rat innervates only the left ovary (Morán et al., 2009), then a negation of the connection in the opposite side can be added to the knowledge model. However, we have yet to implement such modeling for SCKAN. We currently include a textual alert note as an annotation property for such cases.

2.4.2 FAIR Vocabulary Management

All connections in SCKAN are annotated with the SPARC vocabulary (Surles-Zeigler et al., 2022) to ensure anatomical structures are used consistently across all SPARC systems. The overall strategy for the SPARC vocabulary is to utilize anatomical ontologies already in use by the community to conform to the FAIR principles and to ensure that SCKAN is interoperable with other projects in the Common Fund Data Ecosystem (CFDE, NIH Common Fund, 2021). As SCKAN contains connectivity information derived from multiple species, the pan-species anatomy ontology UBERON (Mungall et al., 2012) forms the "backbone" of its anatomical terminology efforts. Species-specific ontologies such as FMA (Rosse and Mejino Jr, 2003) and EMAPA (Hayamizu et al., 2013) are used to supplement as necessary. While SPARC is committed to reusing anatomical terms from standard community ontologies, we often require additional terms not present in community ontologies. To quickly mint identifiers for new terms, we utilize Interlex, an on-line vocabulary management system (Surles-Zeigler et al., 2022). To ensure the validity of new terms and that they are clearly defined, all proposed anatomical terms for SPARC go through specialized pipelines and a curatorial review before being added to the SPARC vocabulary. To make it easier to manage identifiers in SCKAN and perform QC, we restrict IDs for anatomical structures to UBERON and Interlex (ILX) IDs. If a term from another ontology such as FMA is required, it is imported into Interlex and an ILX ID mapped to the source ontology ID automatically generated.

It is critical that all new anatomical terms have the necessary semantics to support queries and linkages required for effective query of SCKAN. All anatomical structures added to Interlex are related explicitly to UBERON terms to allow combined reasoning over UBERON and the new terms. New anatomical structures added are children of anatomical entity (UBERON:0001096). Nerves and nerve plexi are added as children of nerve (UBERON:0001021) and specific ganglia are children of ganglion (UBERON:0000045). Anatomical structures may also be related to additional UBERON entities through the part of relationship as needed to correctly relate individual connections to higher order anatomy. In total, 328 anatomical terms were added to Interlex for SCKAN. The current version of SCKAN contains 242 distinct anatomical structures from UBERON. The InterLex terms representing general anatomical structures are contributed back to community ontologies like UBERON to enhance their coverage of the central and peripheral nervous system. We have 232 terms from InterLex planned for uploading to UBERON.

2.5 Accessing the Knowledgebase: SCKAN Endpoints

SCKAN is accessible through three different graph database endpoints namely: SciGraph, Stardog, and Blazegraph, each serving distinct purposes (Figure 5). SCKAN is periodically loaded into SciGraph, a Neo4j-based graph database that hosts the SPARC knowledge graph. The SPARC knowledge graph leverages SCKAN to enhance the semantics of SPARC data, models, maps, and simulations with ANS connectivity for the SPARC Portal. Accessing SCKAN via SciCrunch also enables users to write queries using Cypher, a declarative query language for Neo4j databases. Some of the example use cases for the SciGraph endpoint are querying over the ApiNATOMY representation and providing terminology services for anatomical annotation. Additionally, SCKAN is available through Stardog, a high-performance graph database equipped with a web interface called Stardog Studio. This interface streamlines the process of testing and editing SPARQL queries. We use Stardog primarily for its web interface, which allows storing and managing frequently used SPARQL queries for SCKAN. It provides an easy access for the knowledge engineers to check the contents of the knowledge base and share their queries with other user groups to expedite their interaction with SCKAN. Lastly, SCKAN can be accessed through a public Blazegraph endpoint, which does not require any registration or authentication. Blazegraph is an ultra-high-performance

graph database that supports RDF/SPARQL APIs. We also use Blazegraph for its fast response time in executing complex SPARQL queries, which might otherwise take several minutes with other graph database systems.

Links for Accessing SCKAN

- SciGraph Endpoint: Link to the GitBook document
- Stardog Endpoint: Link to the GitBook document
- SCKAN Example Queries via Blazegraph: Link to the GitHub document

2.5.1 Making SCKAN Query-Friendly

SCKAN is a sophisticated knowledge base containing a broad collection of complex phenotypic specifications for its neuron populations represented in OWL-DL. These formal specifications allow SCKAN to be verified for consistency, enabling the inference of implicit connectivity knowledge, and classification of its populations through automated reasoning. However, SCKAN's formal representation poses practical challenges in providing a flexible query mechanism for its non-technical end users. In the past, it was challenging for the users like SPARC investigators and anatomical experts to access and retrieve the connectivity knowledge and verify their accuracy. Retrieving this information required writing complex, tedious queries with specific details, understandable only by knowledge engineers trained in formal modeling of ontologies using OWL-DL.

There was not enough query language support that one could comfortably use against the OWL ontologies. The standard graph-based query languages like SPARQL and CYPHER are designed for writing data or instance-level queries against knowledge bases. However, they are not as suitable for writing class-level queries with complex ontological axioms. While it is possible to write SPARQL or CYPHER queries to retrieve information from an OWL ontology, it usually requires an extensive understanding about the domain of the ontology as well as expertise in the technical jargon of OWL-DL. To mitigate this challenge, we developed Simple SCKAN, a query-friendly layer on top of SCKAN.

Simple SCKAN provides an intuitive, higher-level abstraction layer on top of SCKAN's formal OWL axioms. This allows writing simpler queries for SCKAN without any complex OWL-DL jargon. Simple SCKAN encapsulates SCKAN's complex axioms by constructing a set of 'shortcut relations'. We utilized SPRAQL CONSTRUCT queries to transform the complex OWL statements from SCKAN into simpler RDF graph patterns using the shortcut relations. After each SCKAN release, we run a python script which automates the transformation process and loads simple sckan into the Stardog database. To write SPARQL queries for SCKAN, we simply need to know about the relational predicates supported by Simple SCKAN. Refer to the Simple SCKAN documentation to know more about the Simple SCKAN predicates along with example queries. Simple SCAKN is available to query SCKAN via the Stardog and Blazegraph endpoints. However, Simple SCKAN is not available via the SciGraph endpoint.

The code segments in Listing 3 and 4 provides a comparative view of querying SCKAN with and without Simple SCKAN. Both queries return the same results for 'Give me the list of neuron populations in SCKAN along with their origins and destinations'. However, comparing queries clearly demonstrates how the SPARQL queries for SCKAN can become dramatically simpler to write using Simple SCKAN. The example also demonstrates how the Simple SCKAN version of the query is easier to comprehend and more manageable to write, test, and edit. The 24 lines of complex code fragment in Listing 3 (Line 5-25) can effectively be replaced by 4 lines of code in Listing 3 (Line 4-7) using Simple SCKAN.

```
## This is a SPARQL query example without using Simple SCKAN.
SELECT DISTINCT ?Neuron_ID ?Origin ?Destination
WHERE
{
    VALUES (?npo_soma) {(ilxtr:hasSomaLocatedIn)}
    VALUES (?npo_terminal)
      {
        (ilxtr:hasAxonPresynapticElementIn)
        (ilxtr:hasAxonSensorySubcellularElementIn)
      }
    ?Neuron_ID owl:equivalentClass [rdf:type owl:Class ;
                                    owl:intersectionOf ?bn0] .
    ?bn0 rdf:rest*/rdf:first ilxtr:NeuronApinatSimple .
    ?bn0 rdf:rest*/rdf:first [rdf:type owl:Restriction ;
                              owl:onProperty ?npo_soma ;
                              owl:someValuesFrom [a owl:Restriction ;
                                                  owl:onProperty partOf: ;
                                                  owl:someValuesFrom ?Origin_ID]
                             ] .
    ?bn0 rdf:rest*/rdf:first [rdf:type owl:Restriction ;
                              owl:onProperty ?npo_terminal ;
                              owl:someValuesFrom [a owl:Restriction ;
                                                  owl:onProperty partOf: ;
                                                  owl:someValuesFrom ?Destination_ID]
                             ] .
    ?Origin_ID rdfs:label ?Origin .
    ?Destination_ID rdfs:label ?Destination .
}
```

Listing 3. An example SPARQL query without using Simple SCKAN.

```
## This is a SPARQL query example using Simple SCKAN predicates.
SELECT DISTINCT ?Neuron_ID ?Origin ?Destination
{
    ?Neuron_ID rdfs:subClassOf ilxtr:NeuronApinatSimple;
               ilxtr:hasSomaLocation ?Origin_ID;
               ilxtr:hasAxonTerminalLocation |
               ilxtr:hasAxonSensoryLocation ?Destination_ID.

   ?Origin_ID rdfs:label ?Origin.
   ?Destination_ID rdfs:label ?Destination.
}
ORDER BY ?Neuron_ID ?Origin ?Destination
```

Listing 4. An example SPARQL query using Simple SCKAN predicates.

2.5.2 SCKAN Explorer: A Web-Base Query Interface for SCKAN

To make SCKAN more accessible, we developed SCKAN Explorer (`https://services.scicrunch.io/sckan/explorer/`), an intuitive query interface to explore SCKAN's connectivity information based on Simple SCKAN. It was developed as a prototype query interface for the SPARC anatomical experts and knowledge curators to rapidly check and verify existing SCKAN connections without the hurdle of writing intricate SPARQL queries. SCKAN Explorer allows exploring individual neuron populations through a set of auto-complete enabled filters. It also allows visualizing the connectivity

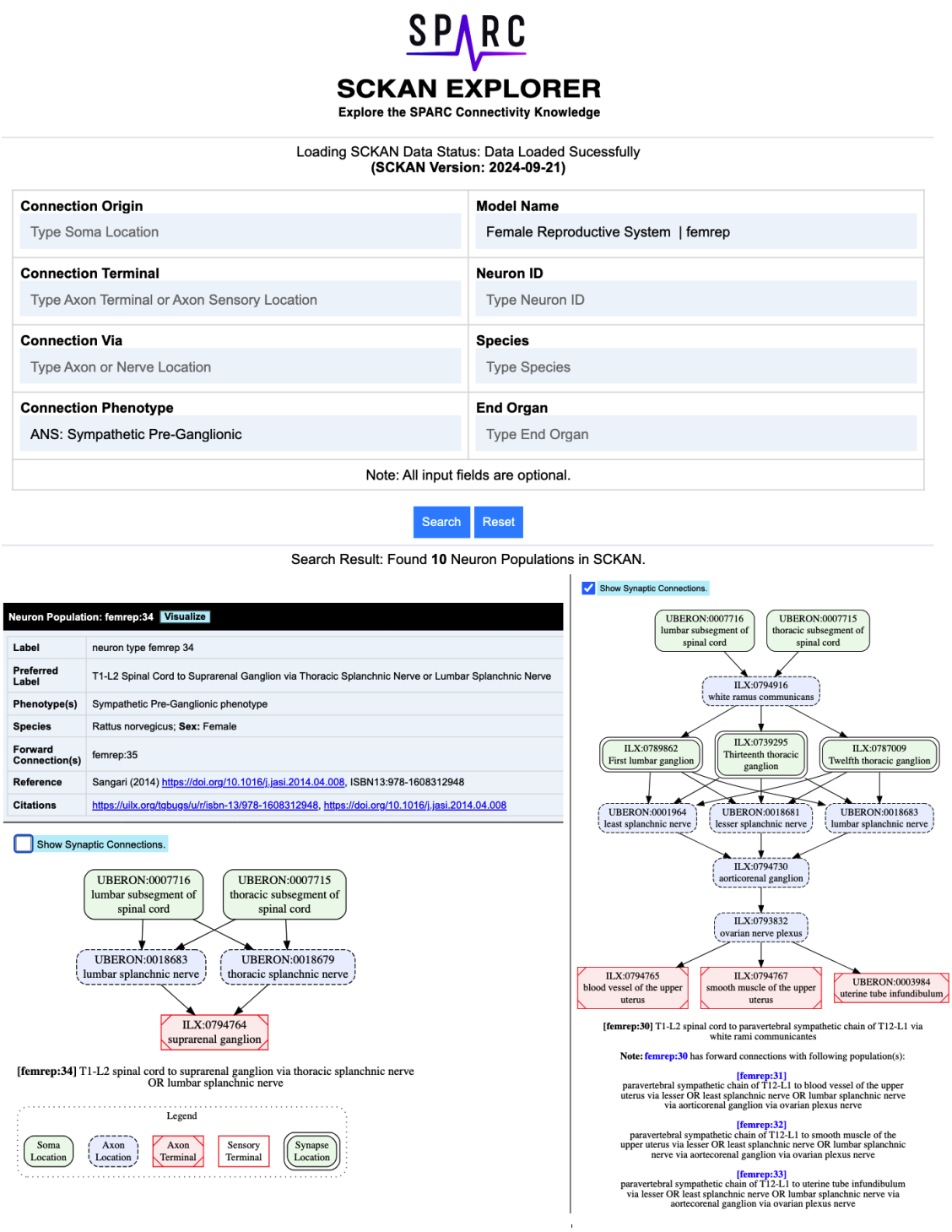

Figure 8. Screenshots from the SCKAN Explorer interface.

pathways of individual neuron populations along with their synapses. A user can search for specific neuron populations based on a given origin (soma location), destination (axon terminal or axon sensory terminal location), and/or the via location (axon location including nerve, nerve plexus, or ganglia).

Figure 8 provides a set of screenshots from SCKAN Explorer. In this example, selecting ‘ANS:Sympathetic Pre-Ganglionic‘ as the connection phenotype for Female Reproductive System returns

10 neuron populations from SCKAN.The screenshot at the bottom-left shows an example population along with its annotation properties and pathways. At the bottom-right, we show an example of a pathway diagram for another sympathetic preganglionic population including its synaptic forward connections. Visualizing the connections in this way can help curators and the domain experts to rapidly interrogate SCKAN to verify the accuracy of its contents. Using SCKAN Explorer, curators can check the integrity of their added information through a set of competency queries, testing routing and partonomies. For example, the "End Organ" query traverses the partonomy chain of major organ systems to allow retrieval of all populations that end in one or more subregions. Checking the number of populations returned often reveals gaps in the partonomy chains present in UBERON or new terms added via Interlex. If such errors are detected, they are corrected by adding the missing partonomy chain into Interlex and importing it into SCKAN.

3 RESULTS

3.1 SCKAN Contents: Statistical Summary

SCKAN releases occur on a roughly quarterly basis. The current SCKAN release (Version: 2024-09-21) represents the end of the initial phase of population driven by the SPARC project, covering ANS connectivity for most major organ systems. It includes some sensory and motor connections as well (see Table 3), especially those related to the vagal system, but the latter were not the focus of this version of SCKAN. In this section, we provide an overview of SCKAN's coverage in terms of its specific contents.

Population Category	Population Count	Soma Location	Axon Location	Axon Terminal	Sensory Terminal
Expert-Contributed Models					
Bolser-Lewis Model of Defensive Breathing	**29**	9	37	15	5
Keast Model of Bladder Innervation	**20**	27	57	20	5
SAWG Model of Bronchomotor Control	**6**	19	11	9	0
SAWG Model of the Descending Colon	**18**	11	17	20	4
SAWG Model of the Pancreas	**5**	11	19	5	0
SAWG Model of the Spleen	**5**	8	10	5	1
SAWG Model of the Stomach	**10**	6	8	8	6
UCLA Model of the Heart	**19**	33	35	19	7
Total	**112**				
Literature-Extracted Connections					
Cranial Nerve Connections	**10**	6	12	10	0
Female Reproductive System	**60**	39	26	54	7
Kidney Connections	**27**	53	44	50	9
Liver Connections	**26**	22	20	27	4
Male Reproductive System	**21**	29	9	8	0
Sweat Gland Connections	**8**	28	15	18	0
Uncategorized Connections	**28**	39	44	25	2
Total	**180**				

Table 2. Count of SCKAN neuron populations along with their locational phenotypes

Dimension	Phenotypic Value	Count
Species	Cat	5
	Dog	22
	Guinea Pig	1
	Human	99
	Mouse	11
	Pig	19
	Rat	86
	Unassigned	88
Sex	Female	69
	Male	25
Locational Phenotype	Soma Location	292
	Dendrite Location	42
	Sensory Axon	37
	Axon Terminal	290
	Sensory Terminal	42
	Axon Location	256
Functional Circuit Role	Excitatory	4
	Inhibitory	4
Circuit Role	Intrinsic	4
	Motor	9
	Sensory	40
ANS Phenotype	Enteric	1
	Parasympathetic	77
	Sympathetic	131

Table 3. Counts of SCKAN neuron populations based on different phenotypic dimensions.

Table 2 summarizes the counts of SCKAN neuron populations along with their locational phenotypes categorized by its two principle sources: Expert-Contributed Models and Literature-Extracted Connections. As shown in the table, SCKAN currently contains 112 expert-contributed populations and 180 literature-extracted populations. Table 3 summarizes the counts of SCKAN populations based on phenotypic values categorized by different phenotypic dimensions.

SCKAN's current focus is on mammalian ANS connectivity. As mentioned already, neuron populations that lack species-specific designation in SCKAN should be assumed to be mammalian by default; i.e., the population was observable in at least one unspecified mammalian species. As reported in the table SCKAN currently has 88 such populations with 'unassigned' species. SCKAN explicitly specifies species for a neuron population when there is at least one piece of experimental evidence documented in the published literature. It is important to note that the species-specific counts of populations reported in Table 3 should not be interpreted as mutually exclusive. For instance, while SCKAN currently includes 99 neuron populations associated with humans; some of these populations may also be found in other species. SCKAN contains several populations that are identified as common across multiple species. For example, the 18 neuron populations in the UCLA Model of the Heart as reported in Table 2 represent connectivity

generally present in dogs, pigs, and humans. As mentioned earlier, under our open-world assumption, we do not assert that a connection represented by a neuron population observed in one set of species is absent in others. However, if evidence indicates that a neuron population present in one species is absent in another, we can close the world (at least partially) by explicitly negating the species-specific relationship for the latter. For example, we might denote that 'neuron-x' is observed in 'species x' but not observed in 'species y' or 'species z' as part of the axiom. At present, however, we have not yet identified populations that are not present in a particular species. Neuron populations in SCKAN that do not have a specified biological sex should, by default, be considered applicable to both sexes. Currently, SCKAN includes 69 populations explicitly associated with female anatomy and 25 populations associated with male anatomy, as shown in Table 3.

As previously mentioned, the current focus of SCKAN is on neural connectivity associated with the distinct subdivisions of the ANS. Consequently, the number of populations associated with ANS subdivision phenotypes predominates the current knowledgebase as reported in Table 3. Nevertheless, we do have small numbers of sensory and motor neuron populations in SCKAN, largely due to their inclusion in some of the connectivity models provided by experts. In addition, as the vagus is a major target of SPARC efforts, we include sensory populations arising from the Nodose ganglion, a sensory ganglion that is physically part of the vagus nerve.

3.2 Competency Queries

SCKAN was designed to answer questions relevant to ANS connectivity. To demonstrate the functionality of SCKAN we developed a set of competency queries (CQ). In this section, we work through the CQs and examine the results. For each question, we provide the SPARQL query and a visualization of the query results. All of the SPARQL queries listed in this section are written using Simple SCKAN. Navigate to the SCKAN CQS Repository for the raw query results in CSV format, along with the visual diagrams produced by the SPARQL queries used in this article.

For the first three CQs we used RAWGraphs to produce the visual query results. RawGraphs (`https://github.com/rawgraphs/rawgraphs-app`) is an open-sourced web tool to create customized vector-based visualizations on top of $d3.js$ library. For the fourth CQ, we developed a custom visualizer utilizing GraphViz (`https://graphviz.org/`), an open source graph visualization library to represent structural information as diagrams.

CQ 1: What connections terminate in bladder? What are the origins of those connections and what are the exact parts of the organ the connections terminate. What nerves are involved in those connections?

The SPARQL query (Listing 5) retrieves the required information based on filtering a set of neuron populations in SCKAN with specific connections to the urinary bladder. The query identifies populations with their soma located in one region (Region A), their axon terminal or axon sensory terminal located in another region (Region B), and the axon potentially passing through or being located in a third region (Region C) which must be a nerve. The focus of the query is on neuron populations where Region B must satisfy the constraint of being a part of the bladder. (Will include a link to the query results). The query for all populations that terminate in bladder returns connections contained within 6 distinct neuron populations that have axon terminals or sensory axon terminals in some part of the bladder. The visual query results in Figure 9 shows the origins of these connections (right panel) and the nerves through which they traverse, including bladder nerve, pelvic splanchnic nerve, and hypogastric nerve.

```
## CQ1: What connections terminate in bladder? What are the origins of those connections and what are the exact parts of the organ the connections terminate? What nerves are involved in those connections?
## Query: This query returns the neuron populations with their origin (soma) located in Region A, axon terminal  located in Region B, and axon located in region C which is a nerve. Region B of the neuron populations must terminate in some part of the bladder.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ilxtr: <http://uri.interlex.org/tgbugs/uris/readable/>

## Select distinct neuron populations and their associated regions
SELECT DISTINCT ?Neuron_ID
?A_IRI ?A_Label ?B_IRI ?B_Label ?C_IRI ?C_Label
?Target_Organ_IRI ?Target_Organ_Label
{
    ## Filter to only include neuron populations connected to the urinary bladder.
    FILTER (?Target_Organ_Label = 'urinary bladder').

    ## Filter to only include C (Axon Location) that is a nerve, nerve plaxus or fiber.
    FILTER (?C_Type in ('nerve', 'nerve plexus', 'nerve fiber'))

    ## Only consider neuron populations from SCKAN.
    VALUES ?SCKAN_Neuron { ilxtr:NeuronSparcNlp ilxtr:NeuronApinatSimple }

    ?Neuron_ID rdfs:subClassOf+ ?SCKAN_Neuron;
               ilxtr:hasSomaLocation ?A_IRI;
               ilxtr:hasAxonLocation | ilxtr:hasAxonToSensoryTerminal ?C_IRI;
               ilxtr:hasAxonTerminalLocation | ilxtr:hasAxonSensoryLocation ?B_IRI.

    ?A_IRI rdfs:label ?A_Label. ?B_IRI rdfs:label ?B_Label. ?C_IRI rdfs:label ?C_Label.

    ## Check if region C is a subclass* of a nerve, nerve plexus, fiber.
    ?C_IRI rdfs:subClassOf*/rdfs:label ?C_Type.

    ## Check if Region B is part of the target organ, either directly or transitively.
    ?B_IRI (ilxtr:isPartOf*)/rdfs:label ?Target_Organ_Label.
    ?Target_Organ_IRI rdfs:label ?Target_Organ_Label
}
ORDER BY ?Neuron_ID ?A_Label ?B_Label ?C_Label
```

Listing 5. The SPARQL query for CQ1.

CQ2: What are all origins and destinations of neuron populations traveling through the vagus?

The SPARQL query (Listing 5) retrieves the required information based on filtering a set of neuron populations in SCKAN with specific connections via the vagus nerve or its branches. The query results show that SCKAN currently has 35 such neuron populations arising from the following 7 distinct origins: inferior vagus X ganglion (nodose ganglion), dorsal motor nucleus of vagus nerve, nucleus ambiguus, superior cervical ganglion, left nodose ganglion, right nodose ganglion, and the nucleus of solitary tract. The query returns 43 distinct destination structures across different organs innervated by the vagus nerve or its branches. The query recognizes the following partonomic vagus nerve branches from SCKAN: External and internal branches of superior laryngeal nerve, External and internal branches of inferior laryngeal nerve, superior laryngeal nerve, abdominal branch of vagus nerve, esophageal vagus trunk, vagus X nerve trunk,

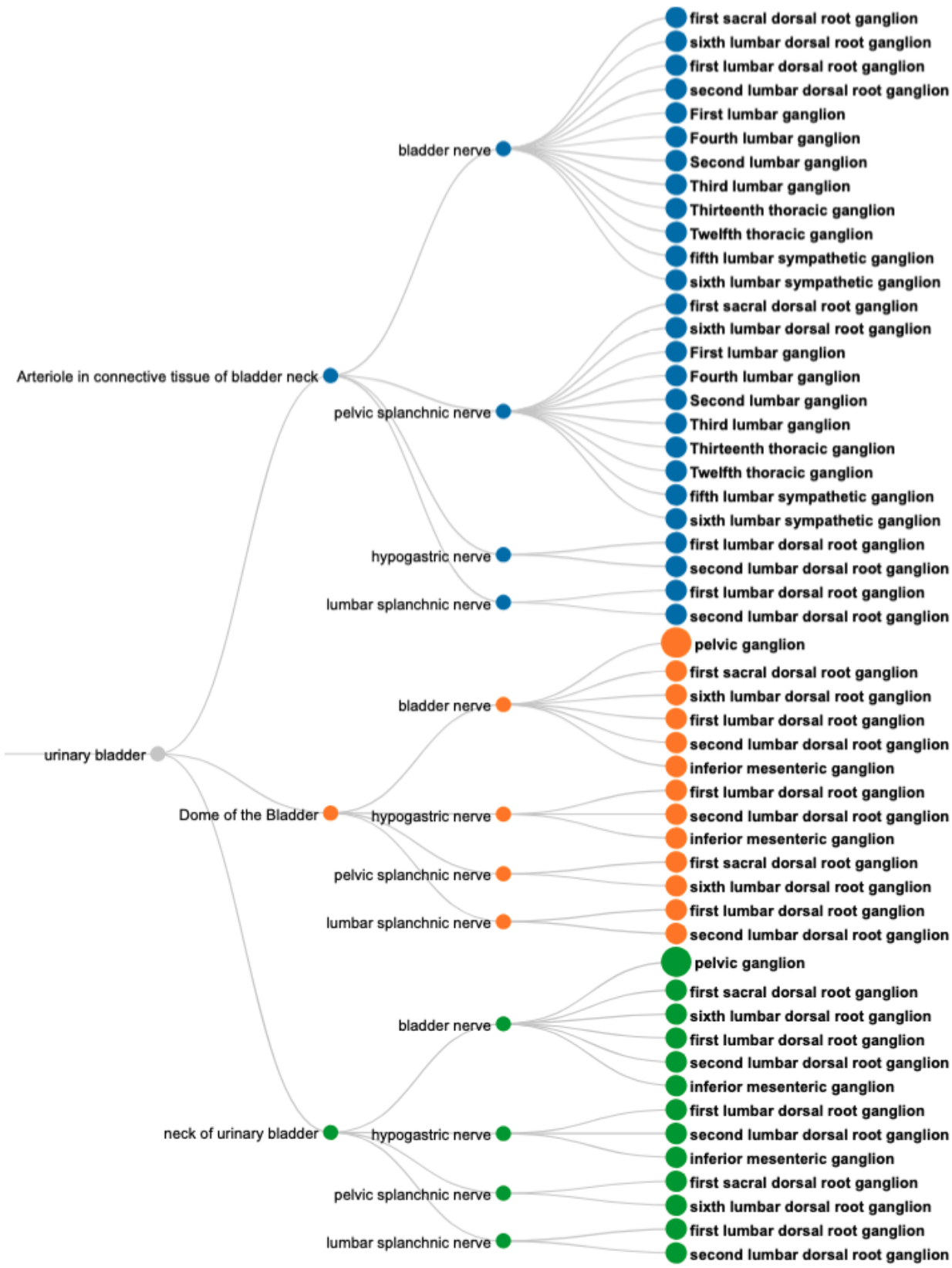

Figure 9. A visual query result for $CQ1$.

anterior and posterior abdominal vagal trunks, and the posterior hepatic branch of vagus nerve. Figure 10 shows the visual results for a portion of connections arising from the inferior vagus X ganglion (nodose ganglion). Please navigate to the full visual result for the remaining connections.

```
# CQ2: What are all origins and destinations of neuron populations traveling through the vagus?
# Query: This query returns any neuron population that has its origin (soma) located in Region A,
#                                axon terminal or axon sensory terminal located in Region B,
#                                and axon located in nerve C where C is vagus nerve or its part.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ilxtr: <http://uri.interlex.org/tgbugs/uris/readable/>

SELECT DISTINCT
?Neuron_ID ?Neuron_Label
?A ?Region_A ?B ?Region_B ?C ?Region_C
{
   FILTER (LCase (?nerve_label) = 'vagus nerve')
   ?Neuron_ID rdfs:label ?Neuron_Label;
              ilxtr:hasSomaLocation ?A;
              ilxtr:hasAxonLocation | ilxtr:hasAxonToSensoryTerminal ?C;
              ilxtr:hasAxonTerminalLocation | ilxtr:hasAxonSensoryLocation ?B.

    ?A rdfs:label ?Region_A. ?B rdfs:label ?Region_B. ?C rdfs:label ?Region_C.

    ## Check if region C is a subclass or partof of the given nerve ether directly or transitively.
    ?C (rdfs:subClassOf* | ilxtr:isPartOf*)/rdfs:label ?nerve_label.
}
ORDER BY ?Neuron_ID ?Region_B ?Region_C
```

Listing 6. The SPARQL query for $CQ2$

CQ3: What anatomical structures might be affected by perturbation of the inferior mesenteric ganglion (IMG)?

This type of query represents one of the driving motivations for SPARC and SCKAN, the ability to better predict on and off target effects for a given electrode placement. Based on the anatomical information in SCKAN, we can pose this query by retrieving all populations that have any part (soma, axon, axon terminal) in the IMG and then retrieving their origins and destinations. The SPARQL query for CQ3 is provided in Listing 7. Figure 11 shows the 55 unique structures that could potentially be affected through a perturbation of the IMG.

```
# CQ3: What anatomical structures might be affected by perturbation of the IMG?
# Query: This query returns the distinct set of anatomical structures connected to the IMG
# based on a set of neuron populations in SCKAN with processes in IMG.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ilxtr: <http://uri.interlex.org/tgbugs/uris/readable/>

SELECT Distinct #?Neuron_ID
?Location_ID ?Location_Label
?Connected_Region_ID ?Connected_Region
{
    FILTER (LCase(?Location_Label) = 'inferior mesenteric ganglion').

    # ilxtr:hasConnectedLocation is equivalent to the implied hasAxonLocation which includes
```

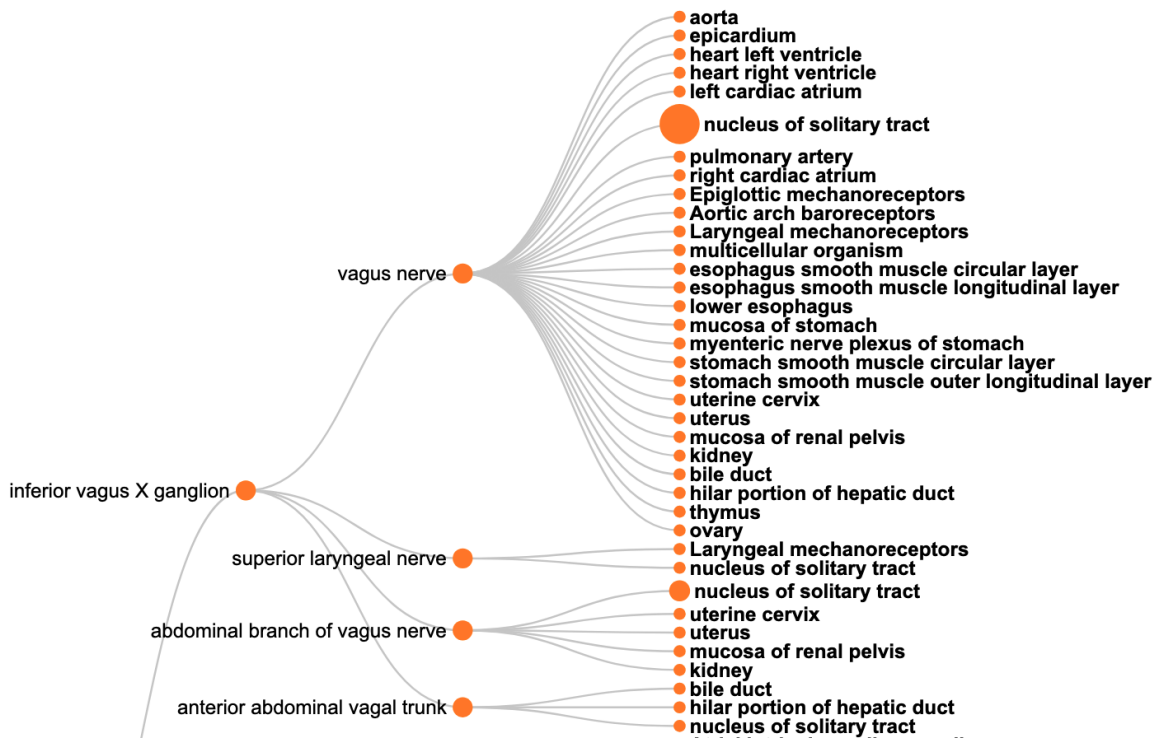

Figure 10. A visual query result for $CQ2$.

```
    # the locations of all parts of a neuron i.e., the dendrite, soma, axon, axon terminal,
    # and axon sensory terminal.
    ?Neuron_ID ilxtr:hasConnectedLocation ?Location_ID.
    ?Location_ID rdfs:label ?Location_Label.

    ?Neuron_ID ilxtr:hasConnectedLocation ?Connected_Region_ID.
    ?Connected_Region_ID rdfs:label ?Connected_Region.

    FILTER (?Location_ID != ?Connected_Region_ID)
}
ORDER BY ?Connected_Region
```

Listing 7. The SPARQL query for CQ3

CQ4: How is sympathetic innervation supplied to the ovaries? Show the origins, terminations, and routing for both pre- and post-ganglionic connections.

This query looks for all pre- and post-ganglionic sympathetic populations connected via a forward connection where the post- ganglionic neurons terminate in the ovary. Neither of the current GUI's make it easy to answer this question. A query through SCKAN Explorer can return all post-ganglionics by entering "ovary" in end organ (both sympathetic and parasympathetic), but retrieving pre-ganglionics requires knowledge of the specific pre-ganglionic populations. SCKAN currently connects pre- and post- ganglionic populations through the 'hasForwardConnection' property, but the property is not reciprocal. Listing 8 provides the SPARQL query for $CQ4$. The query result suggests the innervation involves 5 preganglionic and 6 postganglionic sympathetic neuron populations linked by the synaptic forward connections. Figure 12 shows the visualization of the result. The figure shows the origins, terminations, and routing for both pre- and post-ganglionic connections along with 5 synaptic locations. It should be noted that the visual

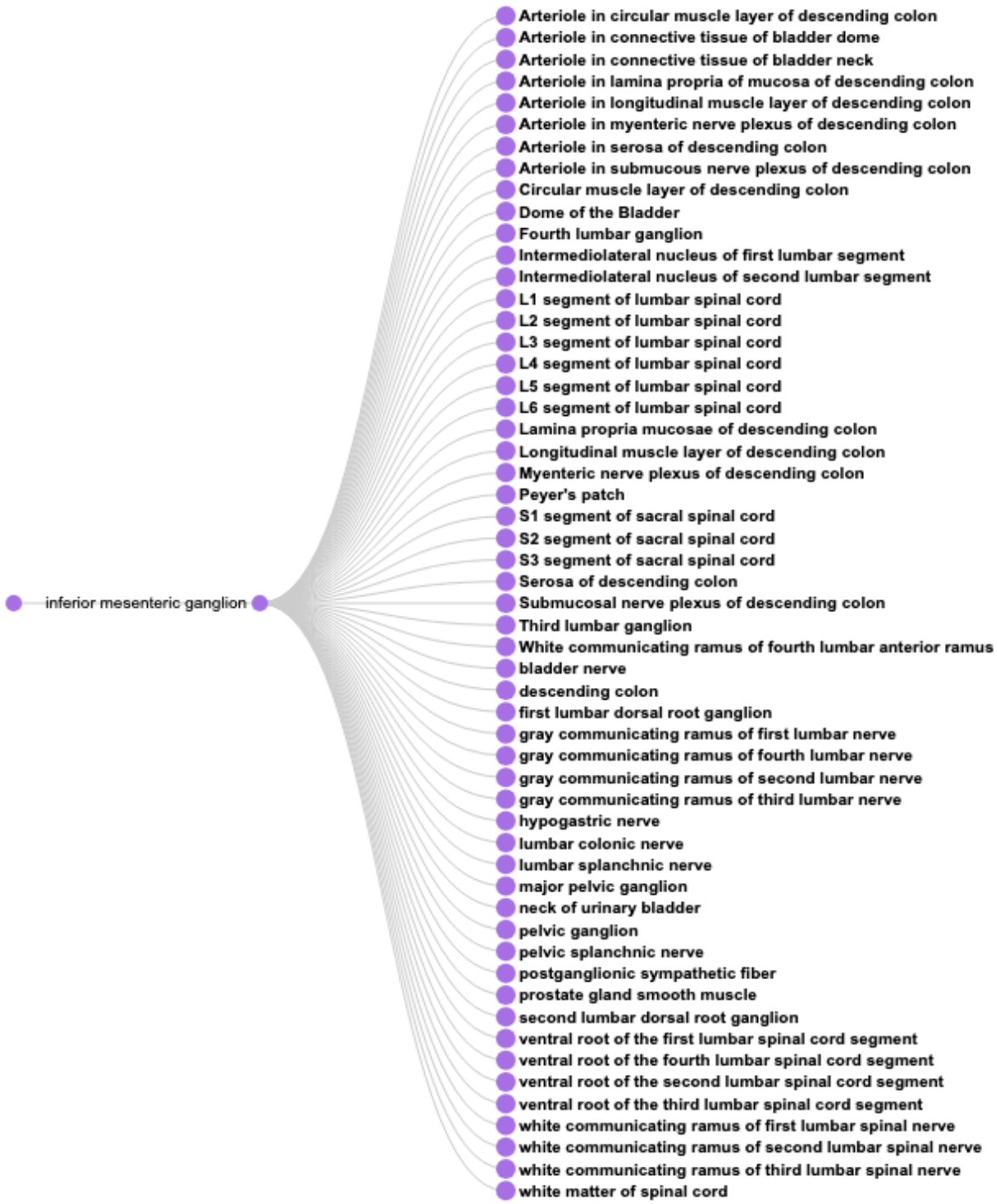

Figure 11. A visual query result for $CQ3$.

query result is simplified for better comprehension by collapsing the nodes representing multiple adjacent segmental structures into single ones (e.g., using a single node like 'T10-L1 segments of spinal cord' instead of having 4 different nodes representing each individual segment between T10 and L1). This is an example of a query where the raw query results cannot be comprehended effectively without the use of

visual diagrams. Visualizing this complex query result demonstrates how SCKAN is laying the foundations for modeling and simulations of neuromodulation therapies.

```
## CQ4: How is sympathetic innervation supplied to an end organ? Show the origins,
## terminations and routing for both pre- and post-ganglionic connections.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ilxtr: <http://uri.interlex.org/tgbugs/uris/readable/>

SELECT DISTINCT ?Neuron_Connected ?V1_ID ?V1 ?V1_Type ?V2_ID ?V2 ?V2_Type ?IsSynapse
WHERE
{
    # Specify the Target_Organ
    FILTER (?Target_Organ = 'ovary')

    # Specify the ANS phenotype for the Target_Neuron
    FILTER (?Target_Neuron_Phenotype = 'Sympathetic Post-Ganglionic phenotype').
    ?Target_Neuron ilxtr:hasNeuronalPhenotype/rdfs:label ?Target_Neuron_Phenotype.

    # Get the Final_Terminal of the Target Neuron.
    ?Target_Neuron (ilxtr:hasAxonTerminalLocation | ilxtr:hasAxonSensoryLocation) ?Final_Terminal.
    # Final terminal of the Target Neuron must be a part of Target Organ.
    ?Final_Terminal (ilxtr:isPartOf* | ilxtr:suppliesTo*)/rdfs:label ?Target_Organ.

    # Get all neurons that are synaptically connected to the Target_Neuron
    OPTIONAL { ?Neuron_Connected ilxtr:hasForwardConnection* ?Target_Neuron. }

    ?V1_ID ilxtr:hasNextNode{ilxtr:isConnectedBy ?Neuron_Connected} ?V2_ID.

    ?V1_ID rdfs:label ?V1. ?V2_ID rdfs:label ?V2.

    # Find the locational phenotypes for the V1 node
    ?Neuron_Connected ?V1_Location_Type_IRI ?V1_ID. ?V1_Location_Type_IRI rdfs:label ?V1_Type.

    # Find the locational phenotype for the V2 node
    ?Neuron_Connected ?V2_Location_Type_IRI ?V2_ID. ?V2_Location_Type_IRI rdfs:label ?V2_Type.

    # Filter out the generic hasConnectedLocation relation for the connected nodes.
    FILTER (ilxtr:hasConnectedLocation not in (?V1_Location_Type_IRI, ?V2_Location_Type_IRI))

    # Check if ?V2_ID is associated with a synapse
    OPTIONAL
    {
        ?Neuron_Connnected ilxtr:hasForwardConnection/ilxtr:hasSomaLocation ?Synapse.
        FILTER (?V2_ID = ?Synapse)
        FILTER (?V2_Type = "hasAxonTerminalLocation")
    }

    # Set IsSynapse to "Yes" if a synapse is detected, otherwise leave it as NO
    BIND (IF (BOUND(?Synapse), "YES", "NO") AS ?IsSynapse)
}
ORDER BY ?Neuron_Connected ?V1 ?V2
```

Listing 8. The SPARQL code for CQ4.

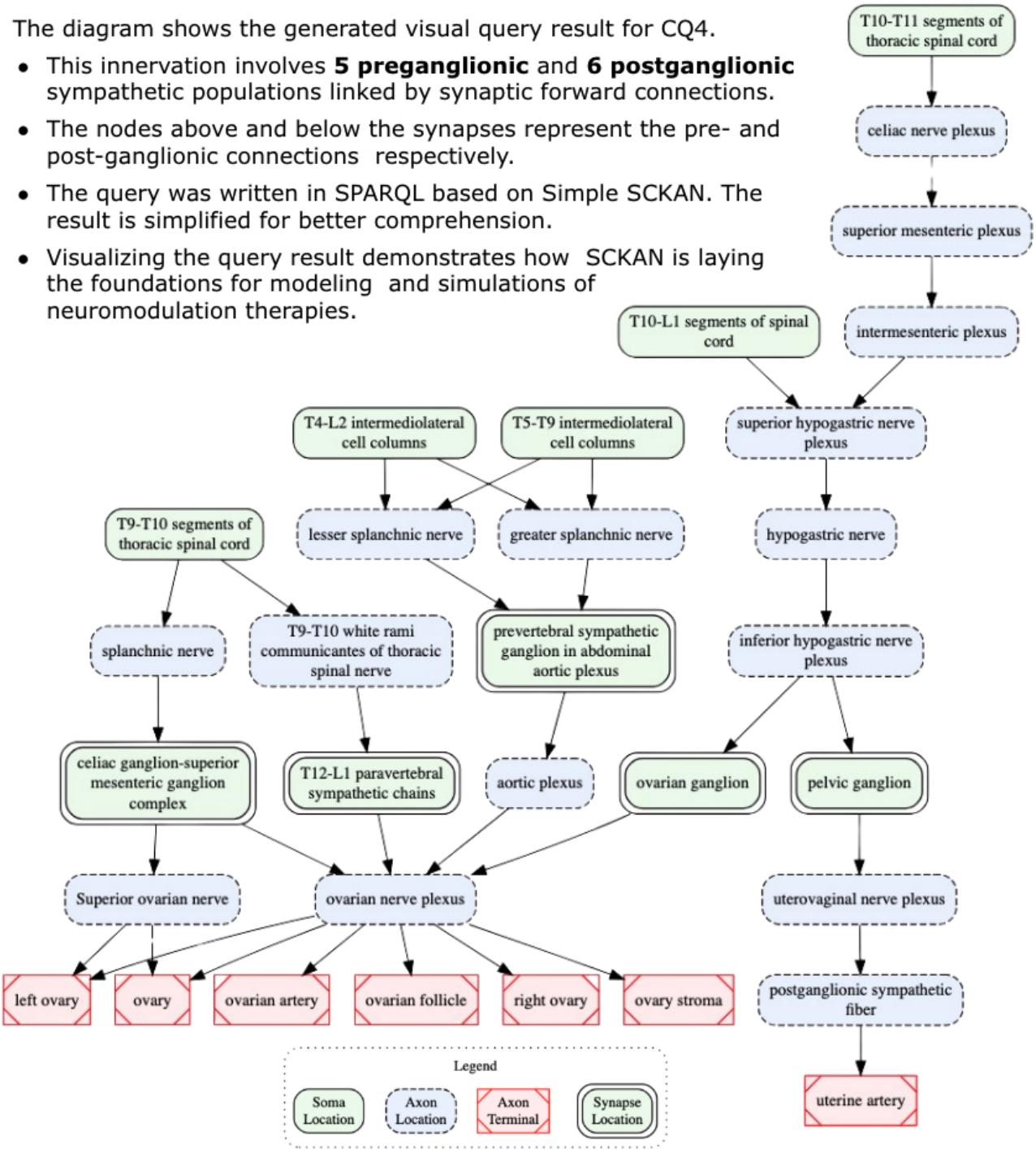

Figure 12. A visual query result for $CQ4$.

Note that the sympathetic innervation of the ovary is an example of an asymmetrical connection, where in the rat, innervation of the left ovary is different than the innervation of the right ovary (Morán et al., 2009). Also note that the sympathetic innervation of the ovary has species specific aspects, as the superior ovarian nerve is apparently not found in humans.

4 DISCUSSION

Neuromodulation therapy and increased attention to the gut-brain axis are shining spotlights on the importance of understanding the anatomical, functional and molecular organization of the peripheral nervous system e.g., e.g., the Allen Mouse Brain Connectivity Atlas (`https://connectivity.brain-map.org/`). To our knowledge, the SCKAN Knowledge Base represents one of the only connectivity knowledge bases that focuses on the peripheral nervous system rather than the central nervous system (ref). Although the current version contains mostly ANS connectivity, reflecting the priorities of the NIH SPARC Program, we are in the process of extending the content with peripheral sensory and motor pathways. The expansion reflects the concomitant expansion of the mission of the SPARC Portal and associated infrastructure as the NIH SPARC Program winds down. SPARC recently became an open data platform for data, models and simulations involving the peripheral nervous system and its end organ, spinal cord, and brain stem interactions. In this way, SPARC complements the infrastructure established through the NIH BRAIN Initiative and similar projects around the world that focus on the CNS. We expect that SCKAN will provide a valuable source of connectivity information to complement the US BRAIN CONNECTS program that is getting underway.

SCKAN was designed to represent anatomical connections in a way that details routing information of axons and locations of somata, axon terminals and axon sensory terminals. In this version of SCKAN, we do not focus on detailed synaptic connections and so dendrites are largely not included. SCKAN connections are automatically visualized via the SPARC 2D and 3D anatomical and functional maps, which are updated with each release of SCKAN on a quarterly basis. The flat maps contain enough detail to understand basic pathways including nerves but retain a level of abstraction to make these routes easy to understand, much like the static images of ANS that are present in resources such as medical textbooks. SCKAN contains more detailed routings which can be returned via queries and visualized in detail. Because SPARC flatmaps are not static drawings but are generated via SCKAN, unlike most atlases, they can be kept up to date as new data are added or edited. We are working on community annotation features whereby users can comment or propose additional connections through the flatmap interface.

To ensure accuracy of the detailed connections in SCKAN, we developed a curatorial workflow that provided multiple checks on connectivity from the initial identification of connections to their coding and query in SCKAN. The information encoded in SCKAN is complex, with assignment of individual parts of a neuron to different locations. Often these different parts of a connection had to be pieced together from multiple scientific articles, especially figures. These parts must then be reassembled in the correct order on to represent a connection. Keeping track of this information and allowing its progressive refinement through the pipeline was very error prone and time consuming when using readily available tools such as spreadsheets. For that reason, we developed a set of tools to support the curators to manage review of sentences extracted from the literature, distillation of knowledge from multiple sources into discrete neuron populations, expression of connections in terms of locations of soma, axon segments and axon terminals followed by mapping to UBERON or Interlex identifiers, and finally coding into OWL. These tools implemented conventions that made it easier for domain experts, curators and knowledge engineers to work together with a shared understanding of what was to be encoded. Once coded it was critical that curators be able to query and visualize the contents of SCKAN for additional QC and to determine whether connections mentioned in the literature were missing.

The curation pipeline involved 3 sets of skills: 1) a curator with neuroscience knowledge sufficient to read and interpret connectivity sources and basic understanding of the SCKAN data model and how to work with ontologies; 2) a domain expert with deep knowledge who could review the information for accuracy

and recommend authoritative sources; 3) a knowledge engineer who could do the deep semantic encoding required. In our case, these skills were embodied in multiple individuals. The curator played the crucial go-between role of identifying populations through translation of the literature and expert knowledge and communicating it in a way that lent itself to knowledge engineering. We initially tried to have a domain expert do this work, but it was difficult for them to express the connections in a form that was easy for the knowledge engineer to encode. Ontological representation is deterministic and not probabilistic, and therefore connectivity statements needed to be expressed in a way that lent themselves to encoding as a set of triples. This type of simplification is often difficult for researchers with deep knowledge of the nuances of connections. We found that constructing single, formulaic sentences in the form of A projects to B via C, each describing a single population, along with relevant parts of papers supporting the population, was the most efficient way for our experts to review them. These sentences also served the critical role as a check throughout the encoding process that what is being expressed in OWL reflects the connection accurately. The new Composer tool allows a curator either to construct a statement manually or automatically by adding the locations of cell soma, axons and axon terminals into a structured form.

4.1 Cell Populations Vs Cell Types

Unlike many connectivity databases, we elected to base connectivity on a neuronal model rather than a regional model. Connections are modeled via the locations of cell bodies, axons and axon terminals which reside in anatomical locations. Axon locations are modeled segmentally so that the routing of an individual axon, which may traverse many anatomical structures and travel via multiple nerves can be accurately traced. Although there is a computational cost to this approach, we elected to employ the neuron populations model for multiple reasons. First, we can more easily accommodate multi-scale connectivity into the same model in that we can provide detailed information about synaptic connectivity at the cell compartment level that then rolls up to regional connectivity statements. Second, as neurons are highly polarized cell types, it allows us to assign molecular and functional phenotypes associated with these populations at a granular level.

Finally, because the cell populations in SCKAN are represented in NPO, they can be compared directly to proposed cell types based on additional phenotypes. The goal of the Neuronal Phenotype Ontology (Gillespie et al., 2022), which forms the backbone of SCKAN, is to be able to represent proposed neuronal types by their phenotypes using a consistent and machine computable representation of molecular, morphological, physiological, anatomical and connectivity phenotypic characters. Using powerful new molecular and correlated techniques, large consortia like the BICCN/BICAN are tackling the open question of defining cell types in the brain (BICCN, 2021). Similar approaches are being employed for the PNS, particularly in the utilization of transcriptomics and spatial transcriptomics to define cell types in peripheral ganglia and cranial nerve nuclei (PRECISION Atlas Paper). The BICCN cell type mapping of the motor cortex suggests that long range axon projection patterns may not map easily onto transcriptomic and epigenetic cell types, suggesting another level of regulation in defining single-cell connectional specificity (BICCN, 2021). Having a consistent model for modeling phenotypes across these multiple dimensions should facilitate understanding of how proposed cell types map onto connectivity patterns in the PNS.

Despite the potential power of utilizing a neuronal model of connectivity, we did encounter some challenges in utilizing the neuron population as the basis of SCKAN. As can be seen in many of the ApiNATOMY diagrams (ref), defining what constitutes a cell population is somewhat arbitrary: do you represent a population that projects from multiple regions to multiple targets as a single population or as multiple cell populations? In our case, we decided to define a population as a population that subserves a

given PNS connection, e.g., the population of pre-ganglionic sympathetic neurons that ultimately enable sympathetic innervation of the heart. So as long as a population could be described as the set of neurons comprising pre-ganglionic sympathetic neurons, the locations of the somata could be distributed. The neuron population approach lends itself to the Flatmap visualization very well, but creating heat maps and other visualizations of query results was challenging, as querying SCKAN Explorer, for example, returns a list of neuron populations rather than providing a regional connectivity matrix. Thus, the user has to do a lot of additional work to find out the answer to a simple query as "What regions project to region X?" As described in a later section, we are developing the next generation of query and visualization tools for SCKAN that will allow users to abstract away the neuron populations to make it easier for users to gain an overview of connectivity for a given query.

4.2 Next Steps

The current version of SCKAN supported by the SPARC program was designed as a resource to aid the neuromodulation community through mapping ANS connectivity derived from existing knowledge and new information coming from SPARC. We decided to focus on well established connections in consultation with ANS experts and not to include every experimental study on ANS connectivity or to include rare variations or minor projections. Nevertheless, variation across connections, especially at the level of nerve branches, is common as we are seeing in early results from SPARC 2. In addition, there are species differences in how connections are routed, and inconsistencies across experimental studies about various aspects of a given connection.

We are addressing these complexities in SCKAN in several ways. At the simplest level, we are using annotation tags like Hippocampome.org to assign a measure of confidence based on the strength of evidence. Second, we are developing a quantitative database for SPARC that can be used to compute quantitative information on distributions of cell somas, terminal fields or nerve trajectories for a given neuron population. For example, our current system for assigning location relies on using segmental labels for spinal cord segments and associated DRGs. However, the spinal cord itself is a single column and it is rare to have populations strictly localized. The bulk of a population is often centered on a few segments with the number of neurons dropping off in either direction. Intersegmental routing of axons is also common, so they do not necessarily only exit at the level at which their soma are located. Finally, minor projections are often noted in certain projections, e.g., projections to the ancillary pelvic ganglia from some organs Keast (1999).

We are also working to provide more detailed provenance for individual parts of a neuron population. In most cases, the detailed locations of soma, axons and axon terminals are constructed from multiple papers. In some cases where we see an inconsistency in a connection between studies, it is often only for part of a population. It would be helpful therefore to be able to annotate individual triples within a given population rather than the entirety of the population. However, current OWL representation requires that annotations be applied to the entire population representation instead of individual parts. We are are planning on moving towards RDF-star Kasenchak et al. (2021) which allows us to annotate individual relations in the RDF graph.

Finally, we are developing new interfaces to make it easier for domain experts to explore and visualize the contents of SCKAN. As can be seen from the competency queries presented in the results, creating an all-encompassing interface that is both easy to use yet encapsulates the complexities of SCKAN is not trivial. To date, we have not found a single visualization that allows easy understanding of these basic types of queries. However, we are in the process of replacing the SCKAN Explorer interface, designed primarily for curators, with a new tool called SCKANNER. SCKANNER will focus on region-to-region

connectivity, abstracting away the individual neuron populations, providing connection matrices, summary reports and the capacity to export search results so that they can be further analyzed. We are also developing new interfaces based on LLM's that will allow researchers to query the knowledge base through natural language queries and be provided with human readable summaries of the results. Both of these tools are currently under development but should be released within the next year.

4.3 Summary and Conclusions

We have described the SCKAN knowledge base, a highly curated knowledge base of detailed connectivity information about the anatomy of the autonomic nervous system. SCKAN is available for query through the SPARC Portal and via a set of programmatic interfaces. SCKAN was designed according to the FAIR principles, utilizing FAIR vocabularies and building from community standards. As efforts like the BRAIN INitiative CONNECTS program start to generate large amounts of CNS activity, we believe it important that we design knowledge bases to be interoperable so that they can be combined to provide a comprehensive, machine-readable representation of nervous system connectivity. The rapid development of neuromodulation therapies and research on the gut-brain axis suggests that we no longer can relegate the PNS to secondary status.

SCKAN was also designed with an eye towards sustainability, through broad-based support and developing automated and semi-automated tools to help with curation and quality control. All the tools developed by SCKAN are available open source. By utilizing an NLP pipeline and a set of tools for extracting and modeling connectivity information, we ensure that information can be kept up to date, while our neuron-population-based data model makes it possible to include both regional and subcellular connectivity and relate connectivity to transcriptomic and other cell type classifications.

Towards this end, SCKAN is currently expanding in scope to support new projects that are utilizing the SPARC tools and services and efforts in the wider ecosystem to map connectivity at the cellular level. To support these projects, SCKAN is increasing its coverage beyond the ANS to provide comprehensive coverage of the entire PNS. As part of its sustainability model, the SPARC DRC has received additional funding from the TARA and NIH HEAL RE-JOIN and NIH PRECISION Human Pain Network projects. TARA is creating an ontology and an on-line 3D atlas of acupoints to serve as a reference for the community, necessitating the addition of nerves and body structures associated with acupoints. The NIH HEAL program is tasked with providing research and solutions to help end long term addiction to opiates. Both RE-JOIN and PRECISION HEAL are looking at cell types and pathways mediating pain.

Finally, with generative AI opening up new possibilities for integrating large swaths of human knowledge, resources like SCKAN become important sources of high quality data. Generative AI like LLMs also open up new possibilities for curation, quality control, query and understanding complex knowledge bases. We expect that such capabilities will make knowledge bases such as SCKAN even more important and useful in our quest to understand the nervous system across scales.

AUTHOR CONTRIBUTIONS

The Author Contributions section is mandatory for all articles, including articles by sole authors. If an appropriate statement is not provided on submission, a standard one will be inserted during the production process. The Author Contributions statement must describe the contributions of individual authors referred to by their initials and, in doing so, all authors agree to be accountable for the content of the work. Please see here for full authorship criteria.

FUNDING

This work is supported by the National Institute of Health (NIH), NIH Common Fund, NIH Office of the Director, Award OT2OD030541.

DISCLOSURES

MEM is a founder and has equity interest in SciCrunch Inc, a tech start up out of UCSD that provides tools and services in support of rigor and reproducibility.

DATA AND CODE AVAILABILITY STATEMENT

- Latest SCKAN Release: Link to the Zenodo repository
- Accessing SCKAN: Link to the documentation
- About Simple SCKAN and Accessing Simple SCKAN: Link to GitHub Repository
- Example Queries in Jupyter Notebook: Link to GitHub Repository
- More About ApINATOMY: Link to the documentation

REFERENCES

de Bono, B., Gillespie, T., Surles-Zeigler, M. C., Kokash, N., Grethe, J. S., and Martone, M. (2022). Representing normal and abnormal physiology as routes of flow in apinatomy. *Frontiers in Physiology* 13. doi:10.3389/fphys.2022.795303

Gillespie, T. H., Tripathy, S. J., Sy, M. F., Martone, M. E., and Hill, S. L. (2022). The neuron phenotype ontology: A fair approach to proposing and classifying neuronal types. *Neuroinformatics* 20, 793

Hayamizu, T. F., Wicks, M. N., Davidson, D. R., Burger, A., Ringwald, M., and Baldock, R. A. (2013). Emap/emapa ontology of mouse developmental anatomy: 2013 update. *Journal of biomedical semantics* 4, 1–5

Kasenchak, B., Lehnert, A., and Loh, G. (2021). Use case: ontologies and rdf-star for knowledge management. In *The Semantic Web: ESWC 2021 Satellite Events: Virtual Event, June 6–10, 2021, Revised Selected Papers 18* (Springer), 254–260

Keast, J. R. (1999). Unusual autonomic ganglia: connections, chemistry, and plasticity of pelvic ganglia. *International review of cytology* 193, 1–69

Kokash, N. and de Bono, B. (2021). Knowledge representation for multi-scale physiology route modeling. *Frontiers in Neuroinformatics* 15. doi:10.3389/fninf.2021.560050

Menke, J., Roelandse, M., Ozyurt, B., Martone, M., and Bandrowski, A. (2020). The rigor and transparency index quality metric for assessing biological and medical science methods. *Iscience* 23

Morán, C., Zarate, F., Morán, J. L., Handal, A., and Domínguez, R. (2009). Lateralization of the connections of the ovary to the celiac ganglia in juvenile rats. *Reproductive Biology and Endocrinology* 7, 1–7

Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E., and Haendel, M. A. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome biology* 13, 1–20

Osanlouy, M., Bandrowski, A., de Bono, B., Brooks, D., Cassarà, A. M., Christie, R., et al. (2021). The sparc drc: Building a resource for the autonomic nervous system community. *Frontiers in Physiology* 12. doi:10.3389/fphys.2021.693735

Ozyurt, I. B., Menke, J., Bandrowski, A., and Martone, M. (2021). Detecting anatomical and functional connectivity relations in biomedical literature via language representation models. In *Proceedings of the Second Workshop on Scholarly Document Processing* (Online: Association for Computational Linguistics), 27–35. doi:10.18653/v1/2021.sdp-1.4

Rosse, C. and Mejino Jr, J. L. (2003). A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of biomedical informatics* 36, 478–500

Surles-Zeigler, M. C., Sincomb, T., Gillespie, T. H., de Bono, B., Bresnahan, J., Mawe, G. M., et al. (2022). Extending and using anatomical vocabularies in the stimulating peripheral activity to relieve conditions project. *Frontiers in Neuroinformatics* 16, 819198

White, T. G., Powell, K., Shah, K. A., Woo, H. H., Narayan, R. K., and Li, C. (2021). Trigeminal nerve control of cerebral blood flow: a brief review. *Frontiers in neuroscience* 15, 649910