

Author: Etienne Doumazane (Nicolas Renier's lab, Paris Brain Institute)

This notebook uses:

- An ontology JSONL file, as downloaded at [https://raw.githubusercontent.com/ChristophKirst/ClearMap2/master/ClearMap/Resources/Atlas/ABA\\_annotation\\_last.jsonl](https://raw.githubusercontent.com/ChristophKirst/ClearMap2/master/ClearMap/Resources/Atlas/ABA_annotation_last.jsonl)
- Cell counts NPY files (1 per sample), as generated with `napari` in a separate notebook.

This notebook generates:

- 2 figures as SVG files
- the cell counts per region as CSV files (1 per sample)

## Imports and function definitions

```
In [ ]: from pathlib import Path
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from natsort import natsorted
```

```
In [ ]: def get_ontology_df_and_dicts(ontology_fpath: str | Path):
    """
    Creates util ontology dataframe and dictionnaires

    ontology_fpath: file path of the ontology file (JSONL file)
    # The ontology file can be downloaded here:
        https://raw.githubusercontent.com/ChristophKirst/ClearMap2/master
    You can use this command in a bash:
        wget https://raw.githubusercontent.com/ChristophKirst/ClearMap2/m
    """

    ontology_df = pd.read_json(ontology_fpath, lines=True)
    dict_colors = dict(zip(ontology_df.name, ontology_df.color_hex_triple))
    dict_names = dict(zip(ontology_df.id, ontology_df.name))
    dict_ids = dict(zip(ontology_df.acronym, ontology_df.id))
    dict_orders = dict(zip(ontology_df.name, ontology_df.allen_brain_inst))

    # optional: metaregion
    metaregions = [dict_ids[acronym] for acronym in ["universe", "root",
    # This line labels all structures with a "metaregion" from the list a
    ontology_df["metaregion"] = ontology_df.structure_path.map(lambda x:
    dict_metaregions = dict(zip(ontology_df.name, ontology_df.metaregion))

    return ontology_df, dict_colors, dict_names, dict_names, dict_orders,
```

```
In [ ]: def get_counts_df(cells_fpath: str | Path):
    """
    cells_fpath: file_path of the cell counts (NPY file)
    Note: needs `dict_colors`, `dict_metaregions`, `dict_orders` at modul
```

```

"""
cells_df = pd.DataFrame(np.load(cells_fpath, allow_pickle=True))
display(cells_df)
counts_df = cells_df.groupby("name").size().sort_values(ascending=False)
counts_df["color"] = counts_df["name"].map(dict_colors)
counts_df["metaregion"] = counts_df["name"].map(dict_metaregions)
counts_df["order"] = counts_df["name"].map(dict_orders)
display(counts_df)
display(counts_df[counts_df.isna().sum(axis=1) > 1])
counts_df = counts_df.dropna()
return counts_df

```

```

In [ ]: ontology_fpath = "~/code/ChristophKirst/ClearMap2/ClearMap/Resources/Atlas
ontology_df, dict_colors, dict_names, dict_names, dict_orders, dict_metar
display(ontology_df)

```

	id	allen_brain_institute_order	acronym	name	color_hex_triplet
0	0	-1	universe	universe	0000
1	997	0	root	root	FFFF
2	8	1	grey	Basic cell groups and regions	BFDA
3	567	2	CH	Cerebrum	B0F0
4	688	3	CTX	Cerebral cortex	B0FF
...	...	...	...	...	...
1323	49	1322	ipf	intraparafloccular fissure	AAAA
1324	57	1323	pms	paramedian sulcus	AAAA
1325	65	1324	pfs	parafloccular sulcus	AAAA
1326	624	1325	IPF	Interpeduncular fossa	AAAA
1327	304325711	1326	retina	retina	7F2E

1328 rows × 13 columns

## 1. Sample 452

```

In [ ]: cells_fpath_452 = Path.home() / "Downloads/452.npy"

```

```

In [ ]: counts_df_452 = get_counts_df(cells_fpath_452)

```

	xt	yt	zt	order	name
0	246.435237	266.958943	24.671975	434	Retrosplenial area, dorsal part, layer 2/3
1	265.029190	285.625289	22.866993	610	Retrosplenial area, dorsal part, layer 5
2	245.965810	233.977027	38.907787	767	Secondary motor area, layer 5
3	259.093759	329.700556	23.968533	610	Retrosplenial area, dorsal part, layer 5
4	260.762971	296.244134	29.414038	610	Retrosplenial area, dorsal part, layer 5
...	...	...	...	...	...
1308	215.433324	204.917844	282.484966	117	optic chiasm
1309	266.213611	311.860260	264.622884	0	universe
1310	392.637415	345.939334	182.518587	382	Field CA1
1311	380.424063	387.477567	109.171312	312782656	No label
1312	241.747543	247.858786	82.056965	687	Retrosplenial area, ventral part, layer 5

1313 rows × 5 columns

	name	count	color	metaregion	order
0	Zona incerta	177	#F2483B	Hypothalamus	802.0
1	Ventral medial nucleus of the thalamus	128	#FF8084	Thalamus	645.0
2	Hypothalamus	112	#E64438	Hypothalamus	715.0
3	Thalamus	104	#FF7080	Thalamus	641.0
4	Lateral hypothalamic area	47	#F2483B	Hypothalamus	794.0
...	...	...	...	...	...
151	optic nerve	1	#CCCCCC	root	1112.0
152	Ventral posteromedial nucleus of the thalamus	1	#FF8084	Thalamus	649.0
153	Anterolateral visual area, layer 5	1	#08858C	Isocortex	168.0
154	Agranular insular area, ventral part, layer 5	1	#219866	Isocortex	294.0
155	ventricular systems	1	#AAAAAA	root	1292.0

156 rows × 5 columns

	name	count	color	metaregion	order
148	No label	1	NaN	NaN	NaN

```
In [ ]: # df = counts_df_452.head(25)
# display(df)
# sns.barplot(data=df, y="name", x="count", hue="name", palette=df["color"]

In [ ]: # df = counts_df_452.sort_values(["order"], ascending=False)
# plt.figure(figsize=(10, 30))
# sns.barplot(data=df, y="name", x="count", hue="name", palette=df["color"]
```

## 2. Sample 450

```
In [ ]: cells_fpath_450 = Path.home() / "Downloads/450.npy"
```

```
In [ ]: counts_df_450 = get_counts_df(cells_fpath_450)
```

	xt	yt	zt	order	name
<b>0</b>	251.134782	337.123348	23.717220	610	Retrosplenial area, dorsal part, layer 5
<b>1</b>	254.127082	335.838686	25.979863	610	Retrosplenial area, dorsal part, layer 5
<b>2</b>	280.289030	312.840724	28.630869	774	Retrosplenial area, lateral agranular part, la...
<b>3</b>	271.906497	285.811063	29.440779	774	Retrosplenial area, lateral agranular part, la...
<b>4</b>	245.084273	251.533330	29.877629	434	Retrosplenial area, dorsal part, layer 2/3
...	...	...	...	...	...
<b>2318</b>	154.310006	244.473793	225.563962	342	Substantia innominata
<b>2319</b>	194.185207	257.731404	225.394987	1097	Hypothalamus
<b>2320</b>	170.077988	233.681987	223.525781	342	Substantia innominata
<b>2321</b>	138.688440	253.334514	217.887971	342	Substantia innominata
<b>2322</b>	194.253224	265.507197	214.115493	797	Zona incerta

2323 rows × 5 columns

	name	count	color	metaregion	order
0	Zona incerta	230	#F2483B	Hypothalamus	802.0
1	Thalamus	178	#FF7080	Thalamus	641.0
2	Hypothalamus	176	#E64438	Hypothalamus	715.0
3	Ventral medial nucleus of the thalamus	160	#FF8084	Thalamus	645.0
4	Lateral hypothalamic area	115	#F2483B	Hypothalamus	794.0
...	...	...	...	...	...
186	Primary somatosensory area, mouth, layer 6a	1	#188064	Isocortex	77.0
187	Primary somatosensory area, trunk, layer 6a	1	#188064	Isocortex	91.0
188	Primary somatosensory area, upper limb, layer 6a	1	#188064	Isocortex	84.0
189	Principal sensory nucleus of the trigeminal	1	#FFAE6F	Pons	889.0
190	Retrosplenial area, dorsal part, layer 2/3	1	#1AA698	Isocortex	327.0

191 rows × 5 columns

	name	count	color	metaregion	order
77	No label	4	NaN	NaN	NaN

```
In [ ]: # df = counts_df_450.head(25)
# display(df)
# sns.barplot(data=df, y="name", x="count", hue="name", palette=df["color"]
```

```
In [ ]: # df = counts_df_450.sort_values(["order"], ascending=False)
# plt.figure(figsize=(10, 30))
# sns.barplot(data=df, y="name", x="count", hue="name", palette=df["color"]
```

### 3. Both sample together

```
In [ ]: df = counts_df_452[["name", "count"]].copy()
df = df.merge(counts_df_452[["name"]], on="name", how="outer").fillna(0)
df["sample"] = "452"

df_2 = counts_df_450[["name", "count"]].copy()
df_2 = df_2.merge(counts_df_450[["name"]], on="name", how="outer").fillna(0)
df_2["sample"] = "450"

df = pd.concat([df, df_2])
df = df.merge(df.groupby("name")["count"].mean().to_frame("mean"), on="name")
df = df.merge(ontology_df[["name", "allen_brain_institute_order", "color"]], on="name")
df["color"] = "#" + df["color"]

display(df)

plt.figure(figsize=(10, 50))
sns.barplot(data=df, y="name", x="count", hue="name", palette=df[["name", "color"]])
sns.stripplot(data=df.query("sample == '452'"), y="name", x="count", hue="sample")
```

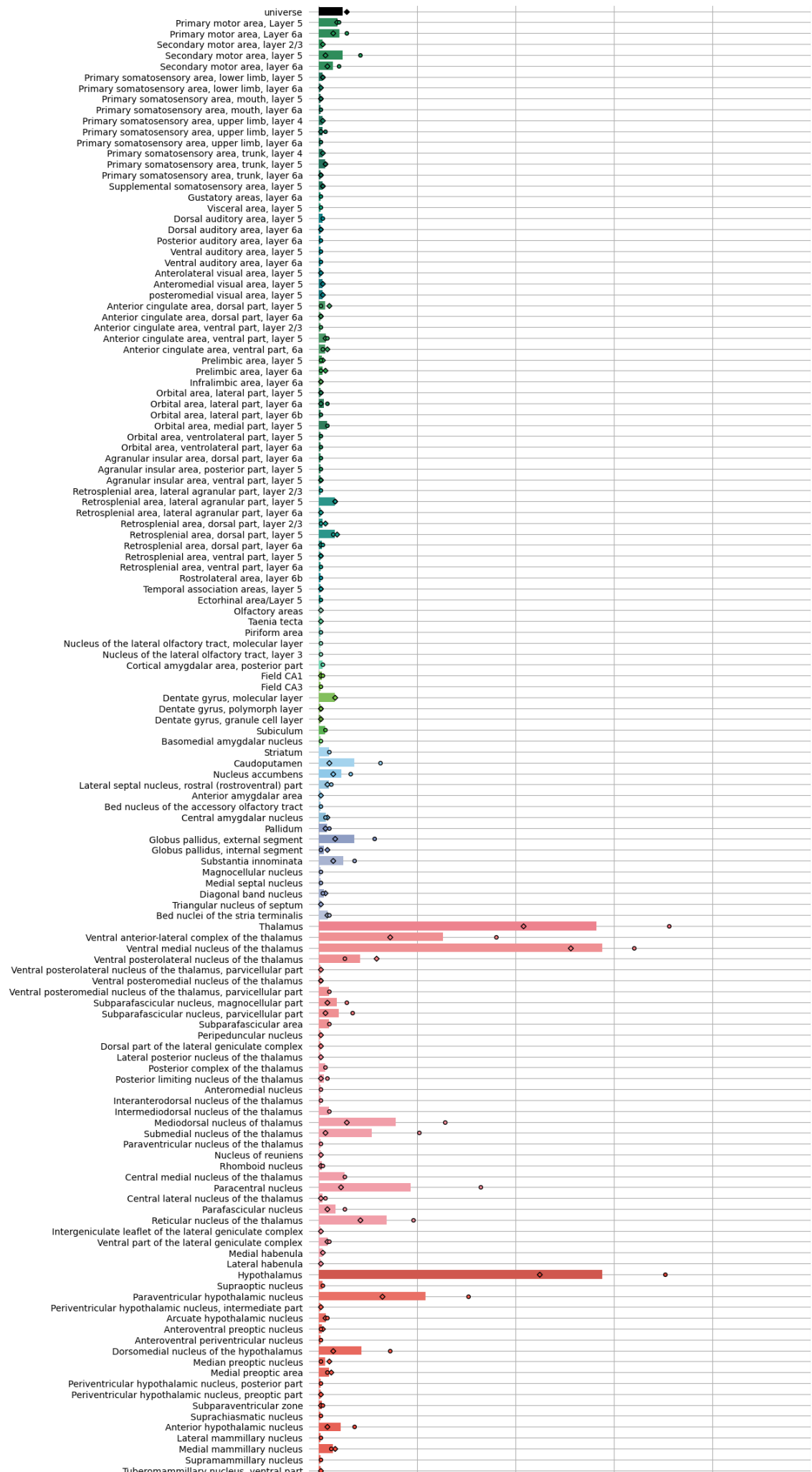
```

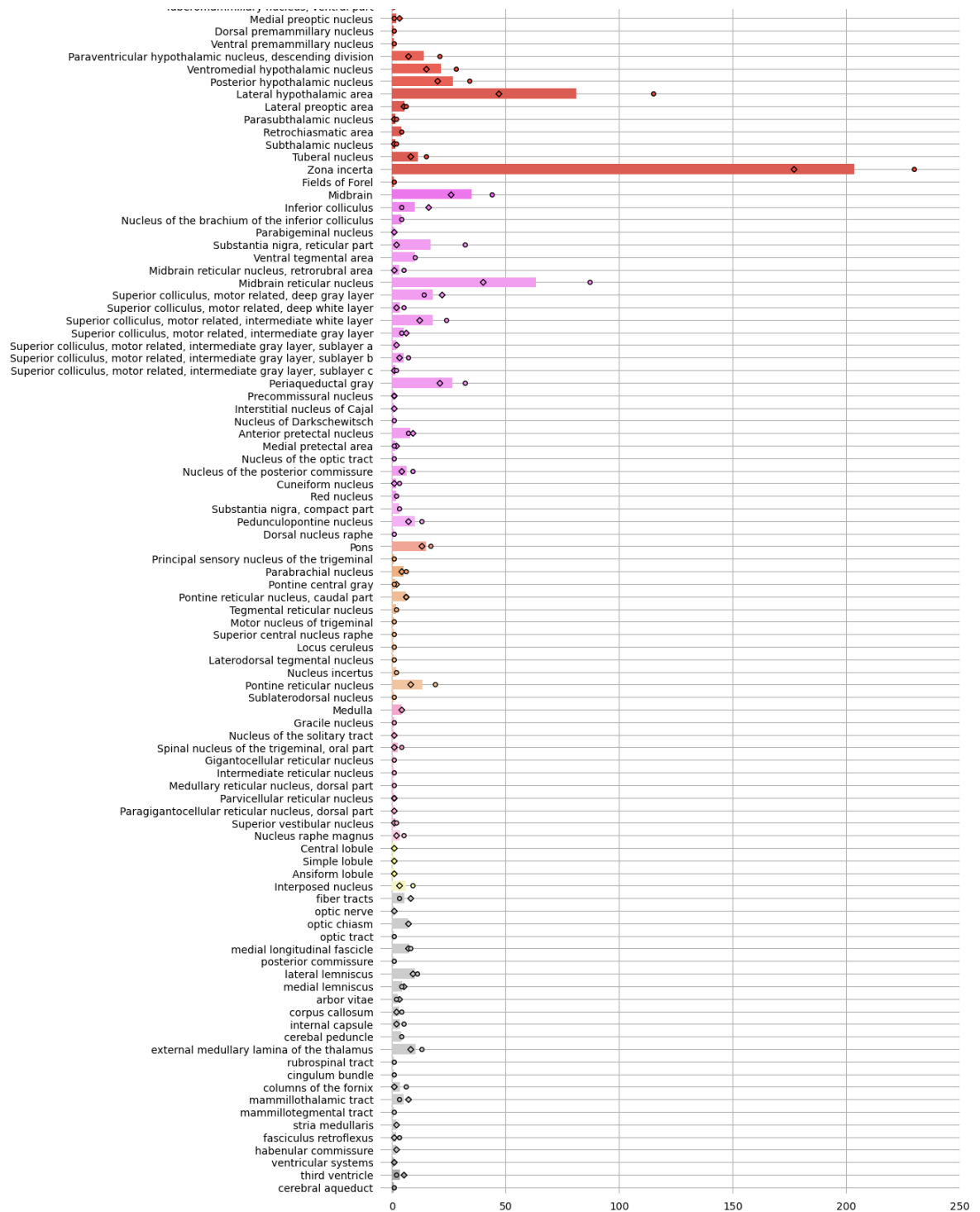
sns.stripplot(data=df.query("sample == '450'"), y="name", x="count", hue=
plt.grid()
plt.xlim(left=-5, right=250)
sns.despine(left=True, bottom=True)
plt.tick_params(bottom=False, left=False)
plt.gca().set_axisbelow(True)
plt.ylabel("")
plt.xlabel("")
plt.savefig("figure.svg", bbox_inches="tight");

```

	name	count	sample	mean	allen_brain_institute_order	color	acronym
<b>0</b>	universe	14	452	12.0	-1	#000000	universe
<b>1</b>	universe	10	450	12.0	-1	#000000	universe
<b>2</b>	Primary motor area, Layer 5	10	450	9.5	21	#1F9D5A	MOp5
<b>3</b>	Primary motor area, Layer 5	9	452	9.5	21	#1F9D5A	MOp5
<b>4</b>	Primary motor area, Layer 6a	7	452	10.5	22	#1F9D5A	MOp6a
...	...	...	...	...	...	...	...
<b>340</b>	habenular commissure	2	452	2.0	1287	#CCCCCC	hbc
<b>341</b>	ventricular systems	1	452	1.0	1292	#AAAAAA	VS
<b>342</b>	third ventricle	2	450	3.5	1299	#AAAAAA	V3
<b>343</b>	third ventricle	5	452	3.5	1299	#AAAAAA	V3
<b>344</b>	cerebral aqueduct	1	450	1.0	1300	#AAAAAA	AQ

345 rows × 7 columns





```
In [ ]: df = counts_df_452[["name", "count"]].copy()
df = df.merge(counts_df_452[["name"]], on="name", how="outer").fillna(0)
df["sample"] = "452"

df_2 = counts_df_450[["name", "count"]].copy()
df_2 = df_2.merge(counts_df_450[["name"]], on="name", how="outer").fillna(0)
df_2["sample"] = "450"

df = pd.concat([df, df_2])
df = df.merge(df.groupby("name")["count"].mean().to_frame("mean"), on="name")
df = df.merge(ontology_df[["name", "allen_brain_institute_order", "color"]], on="name")
df["color"] = "#" + df["color"]
df = df.sort_values("mean", ascending=False).reset_index(drop=True).head(10)

display(df)
```

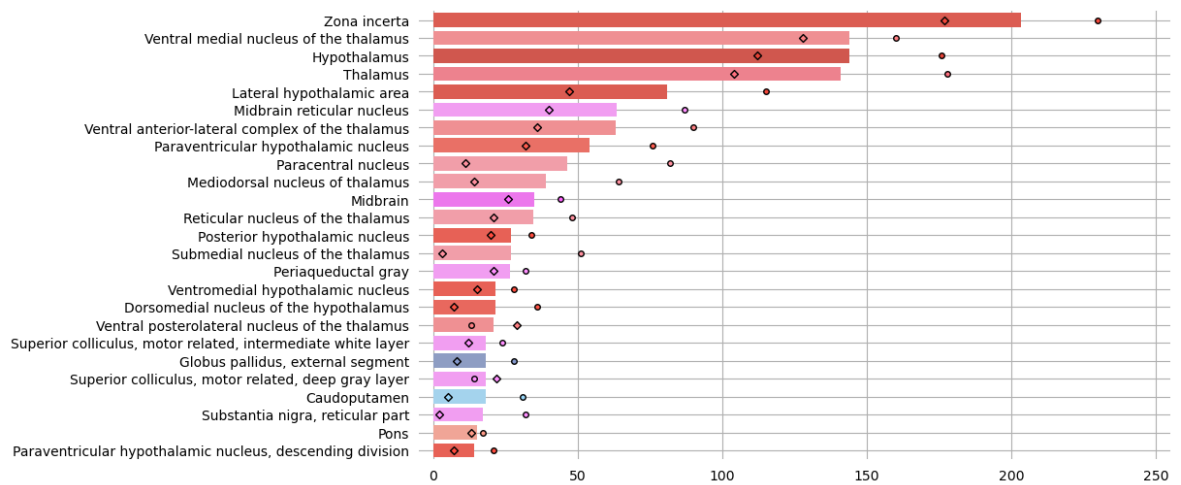


```
plt.figure(figsize=(10, 6))
sns.barplot(data=df, y="name", x="count", hue="name", palette=df[["name",
sns.stripplot(data=df.query("sample == '452'"), y="name", x="count", hue=
sns.stripplot(data=df.query("sample == '450'"), y="name", x="count", hue=
plt.grid()
plt.xlim(left=-5, right=255)
sns.despine(left=True, bottom=True)
plt.tick_params(bottom=False, left=False)
plt.gca().set_axisbelow(True)
plt.ylabel("")
plt.xlabel("")
plt.savefig("figure_2.svg", bbox_inches="tight");
```

	name	count	sample	mean	allen_brain_institute_order	color	acronym
0	Zona incerta	177	452	203.5	802	#F2483B	Z
1	Zona incerta	230	450	203.5	802	#F2483B	Z
2	Ventral medial nucleus of the thalamus	160	450	144.0	645	#FF8084	VM
3	Hypothalamus	176	450	144.0	715	#E64438	H
4	Hypothalamus	112	452	144.0	715	#E64438	H
5	Ventral medial nucleus of the thalamus	128	452	144.0	645	#FF8084	VM
6	Thalamus	178	450	141.0	641	#FF7080	Th
7	Thalamus	104	452	141.0	641	#FF7080	Th
8	Lateral hypothalamic area	115	450	81.0	794	#F2483B	LHA
9	Lateral hypothalamic area	47	452	81.0	794	#F2483B	LHA
10	Midbrain reticular nucleus	40	452	63.5	826	#FF90FF	MRN
11	Midbrain reticular nucleus	87	450	63.5	826	#FF90FF	MRN
12	Ventral anterior-lateral complex of the thalamus	36	452	63.0	644	#FF8084	VAI
13	Ventral anterior-lateral complex of the thalamus	90	450	63.0	644	#FF8084	VAI
14	Paraventricular hypothalamic nucleus	76	450	54.0	720	#FF5D50	PVH
15	Paraventricular hypothalamic nucleus	32	452	54.0	720	#FF5D50	PVH
16	Paracentral nucleus	11	452	46.5	699	#FF909F	PCN
17	Paracentral nucleus	82	450	46.5	699	#FF909F	PCN
18	Mediodorsal nucleus of thalamus	64	450	39.0	685	#FF909F	ME
19	Mediodorsal nucleus of thalamus	14	452	39.0	685	#FF909F	ME

	name	count	sample	mean	allen_brain_institute_order	color	acronym
20	Midbrain	26	452	35.0	806	#FF64FF	ME
21	Midbrain	44	450	35.0	806	#FF64FF	ME
22	Reticular nucleus of the thalamus	48	450	34.5	703	#FF909F	RT
23	Reticular nucleus of the thalamus	21	452	34.5	703	#FF909F	RT
24	Posterior hypothalamic nucleus	34	450	27.0	792	#FF4C3E	PT
25	Submedial nucleus of the thalamus	51	450	27.0	689	#FF909F	SMT
26	Submedial nucleus of the thalamus	3	452	27.0	689	#FF909F	SMT
27	Posterior hypothalamic nucleus	20	452	27.0	792	#FF4C3E	PT
28	Periaqueductal gray	21	452	26.5	838	#FF90FF	PAC
29	Periaqueductal gray	32	450	26.5	838	#FF90FF	PAC
30	Ventromedial hypothalamic nucleus	15	452	21.5	787	#FF4C3E	VMH
31	Ventromedial hypothalamic nucleus	28	450	21.5	787	#FF4C3E	VMH
32	Dorsomedial nucleus of the hypothalamus	7	452	21.5	739	#FF5547	DMH
33	Dorsomedial nucleus of the hypothalamus	36	450	21.5	739	#FF5547	DMH
34	Ventral posterolateral nucleus of the thalamus	13	450	21.0	647	#FF8084	VPI
35	Ventral posterolateral nucleus of the thalamus	29	452	21.0	647	#FF8084	VPI
36	Superior colliculus, motor related, intermedia...	12	452	18.0	833	#FF90FF	SCiv

	name	count	sample	mean	allen_brain_institute_order	color	acronym
37	Globus pallidus, external segment	28	450	18.0	610	#8599CC	GPe
38	Superior colliculus, motor related, deep gray ...	22	452	18.0	831	#FF90FF	SCd <sub>l</sub>
39	Superior colliculus, motor related, intermedia...	24	450	18.0	833	#FF90FF	SCiv
40	Globus pallidus, external segment	8	452	18.0	610	#8599CC	GPe
41	Superior colliculus, motor related, deep gray ...	14	450	18.0	831	#FF90FF	SCd <sub>l</sub>
42	Caudoputamen	31	450	18.0	573	#98D6F9	CF
43	Caudoputamen	5	452	18.0	573	#98D6F9	CF
44	Substantia nigra, reticular part	32	450	17.0	822	#FF90FF	SN
45	Substantia nigra, reticular part	2	452	17.0	822	#FF90FF	SN
46	Pons	13	452	15.0	883	#FF9B88	F
47	Pons	17	450	15.0	883	#FF9B88	F
48	Paraventricular hypothalamic nucleus, descendi...	7	452	14.0	782	#FF4C3E	PVH <sub>l</sub>
49	Paraventricular hypothalamic nucleus, descendi...	21	450	14.0	782	#FF4C3E	PVH <sub>l</sub>



## 4. Count table

```
In [ ]: fname = "counts_by_structure.csv"

df_452 = counts_df_452[["name", "count"]].rename(columns=dict(count="sample_452"))
df_450 = counts_df_450[["name", "count"]].rename(columns=dict(count="sample_450"))

df = df_452.merge(df_450, on="name", how="outer").fillna(0)
df[["sample_452", "sample_450"]] = df[["sample_452", "sample_450"]].astype(int)
df.index = df["name"].map(dict_orders)
df = df.sort_index().reset_index(drop=True)

display(df)
df.to_csv(fname)
```

	name	sample_452	sample_450
0	universe	14	10
1	Primary motor area, Layer 5	9	10
2	Primary motor area, Layer 6a	7	14
3	Secondary motor area, layer 2/3	2	0
4	Secondary motor area, layer 5	3	21
...	...	...	...
224	fasciculus retroflexus	1	3
225	habenular commissure	2	0
226	ventricular systems	1	0
227	third ventricle	5	2
228	cerebral aqueduct	0	1

229 rows × 3 columns