

Using Sankey Diagram to Analyze Drug Pipeline

Tanmay Khole

Orlando, FL

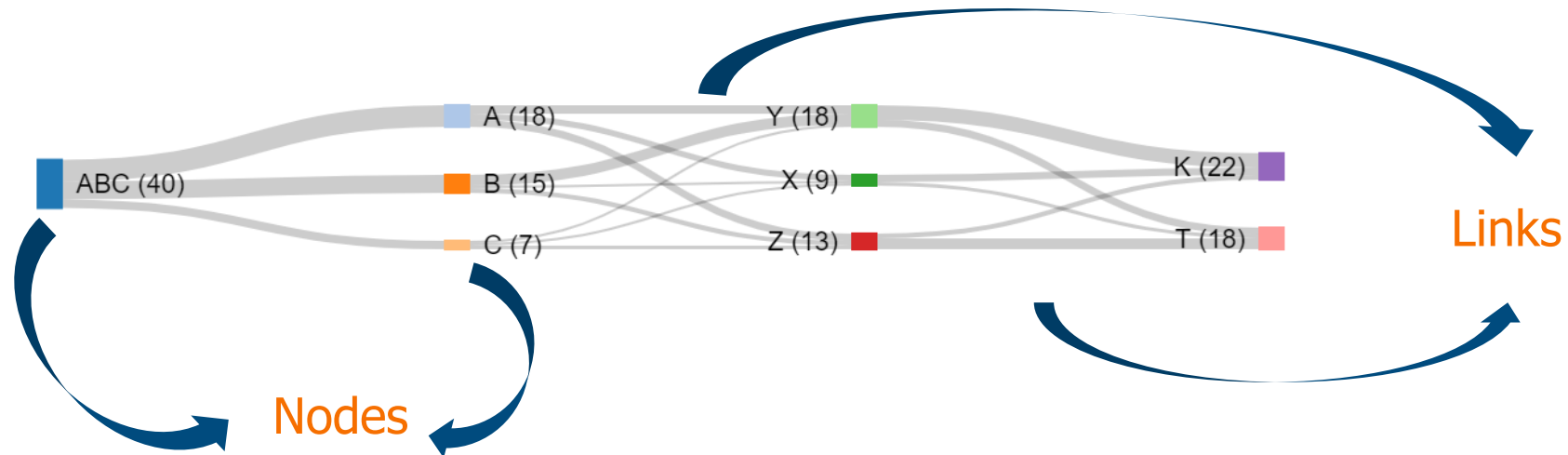
09MAR2020

Disclaimer

- This presentation reflects views of the author and should not be construed to represent any of the clinical trial sponsors' pipeline
- Data analysis and data visualization performed in this paper is not an official representation of any of the sponsor's pipeline but based on the public data acquired from clinicaltrials.gov

Introduction

- A sankey diagram is a visualization used to depict a flow from one set of values to another
- The things being connected are called **nodes** and the connections are called **links**
- Sankeys are best used when you want to show a many-to-many mapping between two domains or multiple paths through a set of stages



Why Sankey Diagram for Drug Pipeline Analysis?

- Each clinical trial sponsor has a unique way of displaying their drug pipeline
- Identifying conditions/compounds focused by the sponsor can be a complex task
- Sankey Diagram of Drug Pipeline:
 - Provides visual representation of the pipeline with respect to its various aspects
 - Facilitates to understand weightage of clinical conditions, interventions, or phases of clinical trials
 - Allows user to see complex pipeline of a sponsor in a single image with a focus on the clinical conditions and interventions/compounds of that sponsor
 - Standardizes the view which helps to compare pipelines of different sponsors

Clinicaltrials.gov Data

- A comprehensive US government database with several datapoints & available for public use
- Below datapoints are considered from clinicaltrials.gov data for data analysis:
 - Lead Sponsor
 - Clinical conditions
 - Interventions
 - Phases of clinical trials

	<i>Sponsor</i>	<i>Distinct On-going* Clinical Trials Count</i>	<i>Data Extraction Date</i>
<i>Sponsor 1</i>	Bristol-Myers Squibb	250	22NOV2019
<i>Sponsor 2</i>	Janssen	126	22JAN2020
<i>Sponsor 3</i>	Merck & Co.	173	
<i>Sponsor 4</i>	Amgen	56	
<i>Sponsor 5</i>	Bayer	56	

**Clinical trials with status: "Active, not recruiting", "Available", "Enrolling by invitation", "Not yet recruiting", or "Recruiting" are considered as on-going*

Steps to Create Sankey Diagram



Data extracted from
clinicaltrials.gov



	RANK	NCT_NUMBER	TITLE	ACRONYM	STATUS	STUDY_RESULTS	CONDITIONS	INTERVENTIONS	OUTCOME_MEASURES	SPONSOR_COLLABOR	GENDER	AGE	PHASES
1	8	NCT01919116	Study of Early	SELECT	Not yet recruit	No Results Available	Relapsed or Refr.	Drug: Carfilzomib/Cr.	Number of part.	Angen	AB	18 Years to 99.	PHASE 2
2	10	NCT04060181	Tafametin LA		Not yet recruit	No Results Available	Metastatic	Drug: Tafametin LA	Objective Resp.	Angen/Beck.	AB	18 Years to 99.	PHASE 2
3	11	NCT04060028	Efficacy and S.		Not yet recruit	No Results Available	Systemic Lupus E.	Drug: JAK1/2 Inhib.	Per cent of pat.	Angen	AB	18 Years to 75.	PHASE 2
4	12	NCT03971071	A Study to Eval.		Recruiting	No Results Available	Migraine Headache	Drug: Erenumab 70.	Number of part.	Angen	AB	18 Years to 65.	PHASE 4
5	13	NCT03963029	A Study of Effic.		Recruiting	No Results Available	Secondary Hyper.	Drug: Elicalcetide	Percent chang.	Angen	AB	2 Years to 11 Y.	PHASE 3
6	14	NCT03937154	Study of Romip.		Not yet recruit	No Results Available	Chemotherapy-in.	Drug: Romiplostin D.	Incidence of ad.	Angen	AB	18 Years to 10.	PHASE 3
7	15	NCT03954684	Study to Evalu.		Recruiting	No Results Available	Relapsed Refract.	Drug: Drug: Carfilz.	Number of part.	Angen	AB	18 Years and o.	PHASE 4
8	16	NCT03927144	Study of Suniti.		Recruiting	No Results Available	Epidemic Migraine	Drug: Erenumab/Cr.	Proportion of s.	Angen/Novartis	AB	18 Years and o.	PHASE 4
9	17	NCT03912337	Effect of Erenu.		Recruiting	No Results Available	Migraine	Drug: Pooled Drug	Sum of month.	Angen/Novartis	AB	18 Years to 99.	PHASE 4
10	19	NCT03975401	Effect of Erenu.	VESALIUS-CV	Recruiting	No Results Available	Coronary Heart Di.	Drug: Evolocumab/D.	Time to Corvira.	Angen	AB	50 Years to 79.	PHASE 3
11	20	NCT03959427	A Study Comp.	APPROVQ	Recruiting	No Results Available	Relapsed or Refr.	Drug: Carfilzomib/Cr.	Overall Respon.	Angen	AB	18 Years to 70.	PHASE 3
12	23	NCT03936040	Efficacy and S.	OASIS (EM)	Recruiting	No Results Available	Migraine	Drug: Erenumab Dos.	Change from b.	Angen	AB	6 Years to 11 Y.	PHASE 3
13	24	NCT03932996	Efficacy and S.	OASIS (CM)	Recruiting	No Results Available	Migraine	Drug: Erenumab Dos.	Change from b.	Angen	AB	6 Years to 11 Y.	PHASE 3
14	26	NCT03819607	A Study Evaluat.	DOHIA	Recruiting	No Results Available	Paroxysmal Noct.	Drug: ADP 958 Drug.	Hemolysis es.	Angen	AB	18 Years and o.	PHASE 3
15	27	NCT03812224	A Controlled Tr.		Active, not rec.	No Results Available	Migraine	Drug: Erenumab	Mean Monthly.	Angen	AB	20 Years to 65.	PHASE 3
16	28	NCT03809663	A Dose Rangen.		Recruiting	No Results Available	Atopic Dermatitis	Drug: Tricafonium	Investigator's	Angen/Novartis	AB	18 Years to 75.	PHASE 2
17	35	NCT03795902	Study to Assess	METFORMIN-HF	Recruiting	No Results Available	Heart Failure With	Drug: Olanzapine-HF	Change in pes.	Angen/Cytobion	AB	18 Years to 65.	PHASE 3
18	37	NCT03638128	Open-label Ext.		Recruiting	No Results Available	Osteogenesis Imp.	Drug: Denosumab	Subject incid.	Angen	AB	5 Years to 20 Y.	PHASE 3
19	38	NCT03633708	Study of Effic.		Recruiting	No Results Available	Secondary Hyper.	Drug: Elicalcetide	Change of vira.	Angen	AB	up to 18 Years.	PHASE 2
20	41	NCT03608883	A Phase 1/2 S.		Recruiting	No Results Available	KRAS p.G12C Mu.	Drug: AMG 510	Number of Part.	Angen	AB	18 Years to 10.	PHASE 2

Data filtered & converted to SAS® data format



Sankey diagram created by
%sankey2html



Data analyzed by
%sankey_nodes

	COND	INT	PHASES	SPNR
1	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
2	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
3	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
4	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
5	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
6	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
7	CARDIOVASCULAR & KIDNEY DISEASE	EVOLUCIMAB	PHASE 3	ANGEN
8	CARDIOVASCULAR & KIDNEY DISEASE	ETELCALCETIDE	PHASE 3	ANGEN
9	CARDIOVASCULAR & KIDNEY DISEASE	ETELCALCETIDE	PHASE 3	ANGEN
10	CARDIOVASCULAR & KIDNEY DISEASE	ONECANTY MECARIL	PHASE 3	ANGEN
11	CARDIOVASCULAR & KIDNEY DISEASE	ONECANTY MECARIL	PHASE 3	ANGEN
12	CARDIOVASCULAR & KIDNEY DISEASE	ROMIPLOSTIN	PHASE 3	ANGEN
13	CARDIOVASCULAR & KIDNEY DISEASE	ROMIPLOSTIN	PHASE 3	ANGEN
14	DIABETES & ENDOCRINOLOGY DISORDERS	DEVOSUMAB	PHASE 3	ANGEN
15	DIABETES & ENDOCRINOLOGY DISORDERS	ETELCALCETIDE	PHASE 3	ANGEN
16	DIABETES & ENDOCRINOLOGY DISORDERS	ETELCALCETIDE	PHASE 3	ANGEN
17	DIABETES & ENDOCRINOLOGY DISORDERS	ETELCALCETIDE	PHASE 3	ANGEN
18	INFLAMMATION & IMMUNOLOGY	TEZPEPELINAB	PHASE 2	ANGEN
19	INFLAMMATION & IMMUNOLOGY	ANG-570	PHASE 2	ANGEN
20	INFLAMMATION & IMMUNOLOGY	ANG-592	PHASE 2	ANGEN
21	INFLAMMATION & IMMUNOLOGY	ANG-592	PHASE 2	ANGEN
22	INFLAMMATION & IMMUNOLOGY	ANG-592	PHASE 2	ANGEN

Horizontal data mapped to vertical data



Data Mapping

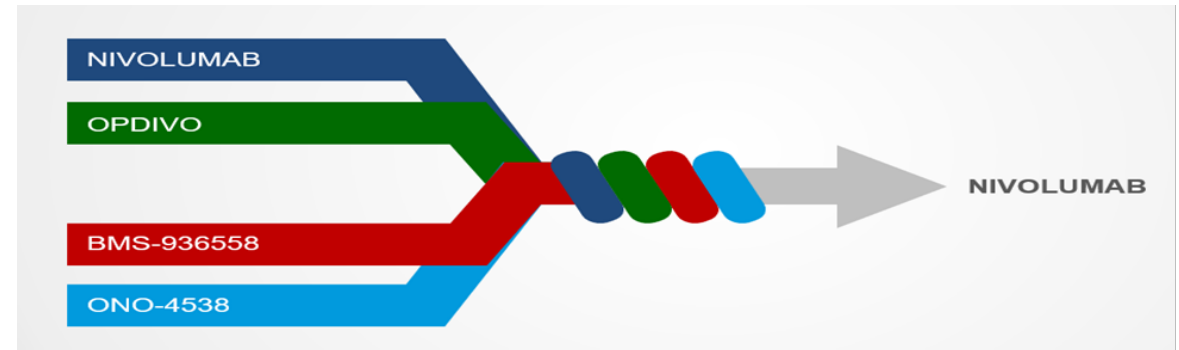
- Data mapping is an essential component in order to connect links and nodes in sankey diagrams
- Clinical trials data obtained from clinicaltrials.gov contain
 - Multiple names for same conditions (e.g.: “NSCLC”, “Non-Small Cell Lung Cancer”, or “Carcinoma, Non-Small-Cell Lung”)
 - Multiple names for same drug/biologic compounds (e.g.: “Nivolumab”, “Opdivo”, “BMS-936558”, “ONO-4538”)
- Important to map similar terms into same category

Data Mapping - Examples

Clinical conditions representing same condition are mapped to same category



Interventions/drugs representing same compound are mapped to same category



Clinical conditions belonging to same category are mapped to high-level category



Data Analysis

- Data analysis is performed by calculating number of objects with respect to its categories which needs to be displayed in sankey diagram
- The categories are used as nodes and the count of those objects are used to determine the width of links between the selected categories
- `%sankey_nodes` will calculate the number of objects, in this case, number of clinical trials

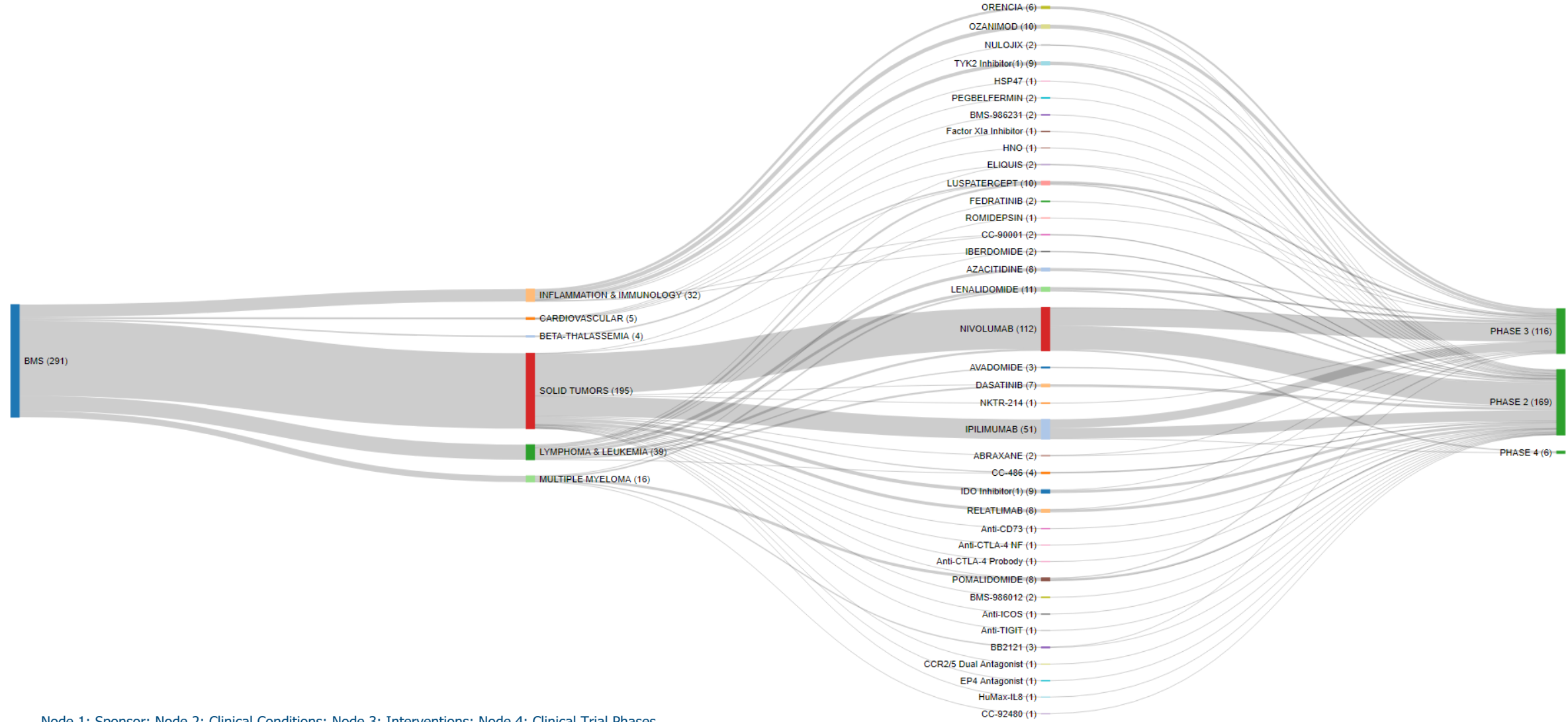
```
%sankey_nodes(inds = ct_gov  
               , outds = sankey_out  
               , nodes=%str(sponsor|conditions|interventions|phases)  
               , cond =  
               );
```

Data Visualization

- Data visualization step is performed using
 - SAS® macro `%sankey2html` and
 - D3.js (JavaScript library)
- The output is created in HTML format

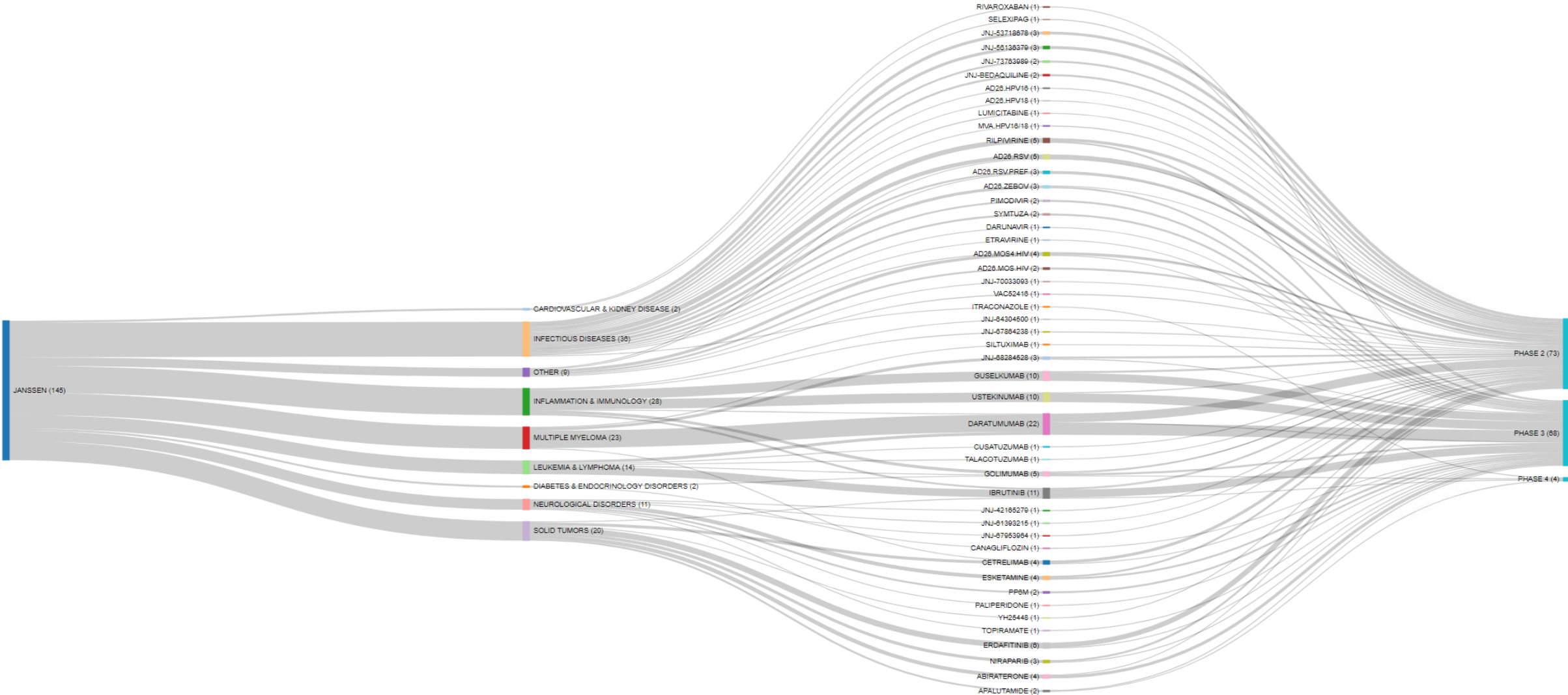
```
%sankey2html(indata = %nrquote(&sankeydata.)  
              ,outfl  = %sysfunc(pathname(outg,f))/sankey.html  
              ,widthn = 2100  
              ,height  = 700  
              ,flow_num =  
              );
```

Sankey Diagram - BMS



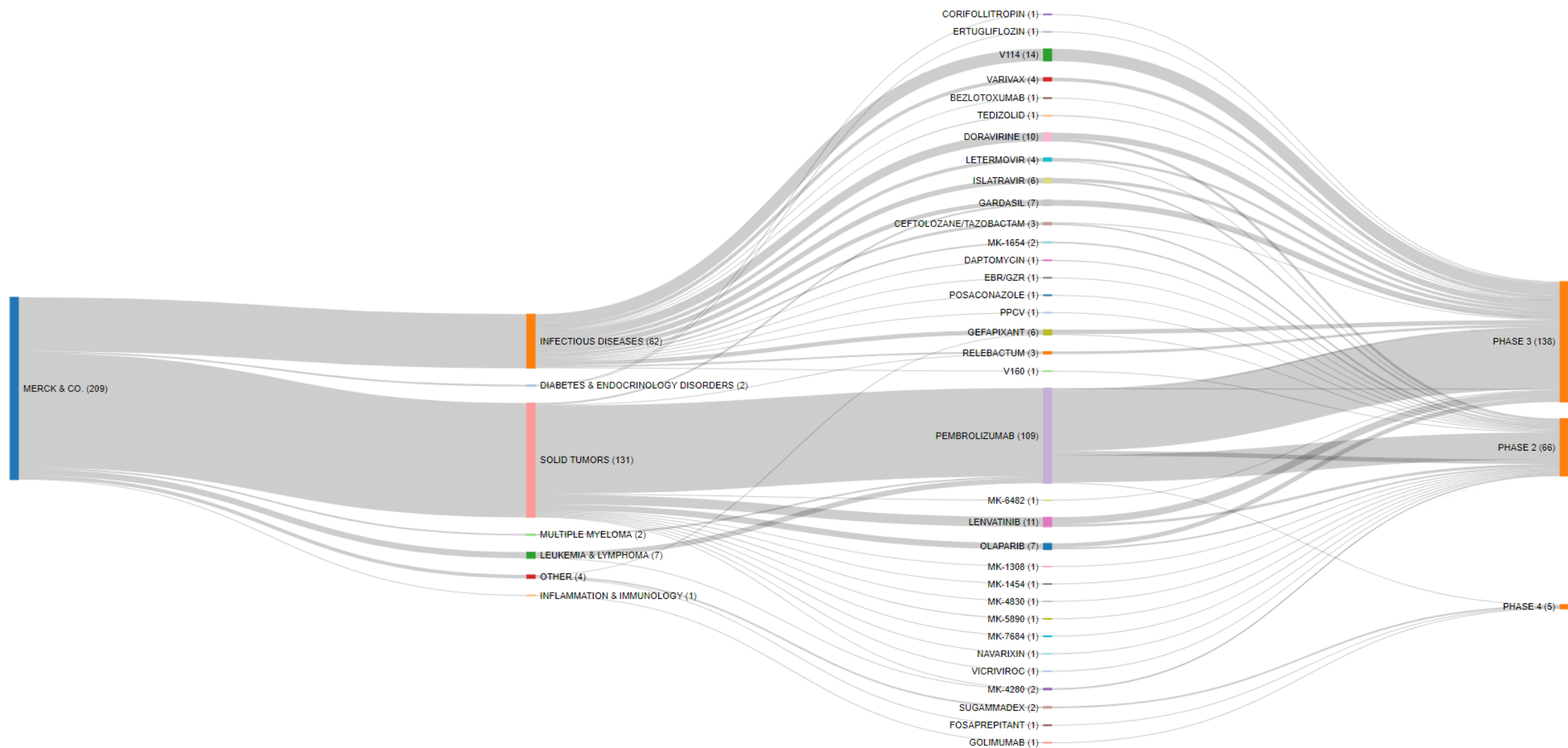
Node 1: Sponsor; Node 2: Clinical Conditions; Node 3: Interventions; Node 4: Clinical Trial Phases
Number of on-going clinical trials for each node are displayed in parenthesis.
Clinical trials with multiple phases are counted toward the higher phase; Clinical trials with multiple clinical conditions are counted towards each condition; Clinical trials with multiple interventions are counted towards each intervention.
Note: Data analysis and data visualization performed in this paper is not an official representation of any of the sponsor's pipeline but based on the data acquired from clinicaltrials.gov.

Sankey Diagram - Janssen



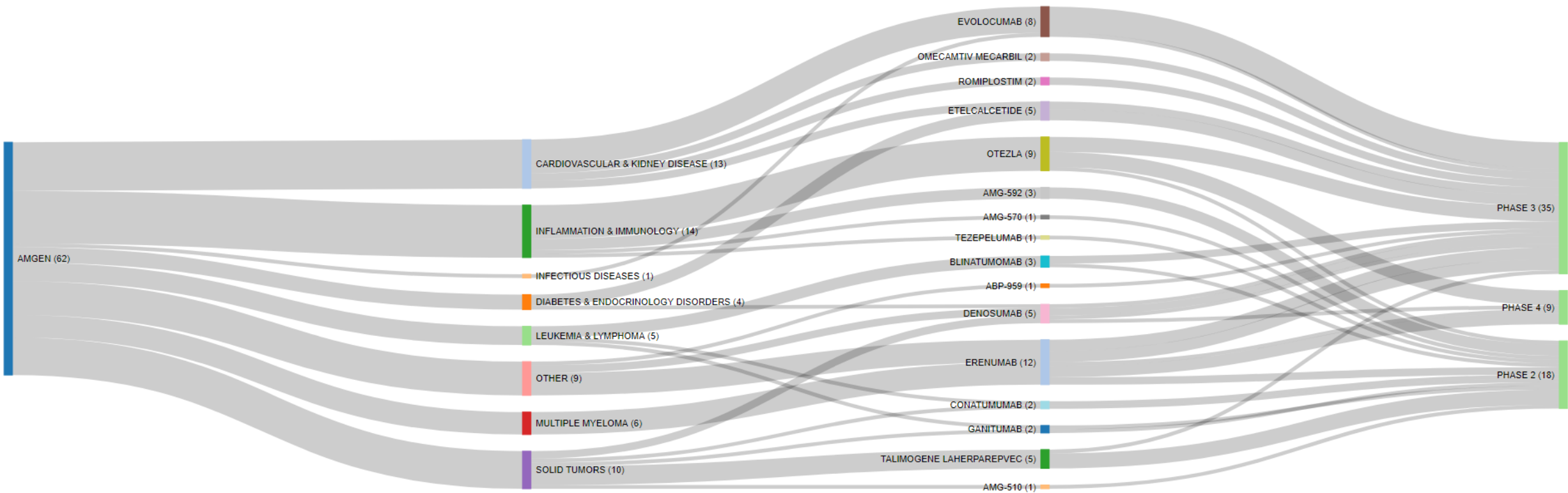
Node 1: Sponsor; Node 2: Clinical Conditions; Node 3: Interventions; Node 4: Clinical Trial Phases
Number of on-going clinical trials for each node are displayed in parenthesis.
Clinical trials with multiple phases are counted toward the higher phase; Clinical trials with multiple clinical conditions are counted towards each condition; Clinical trials with multiple interventions are counted towards each intervention.
Note: Data analysis and data visualization performed in this paper is not an official representation of any of the sponsor's pipeline but based on the data acquired from clinicaltrials.gov.

Sankey Diagram – Merck & Co.



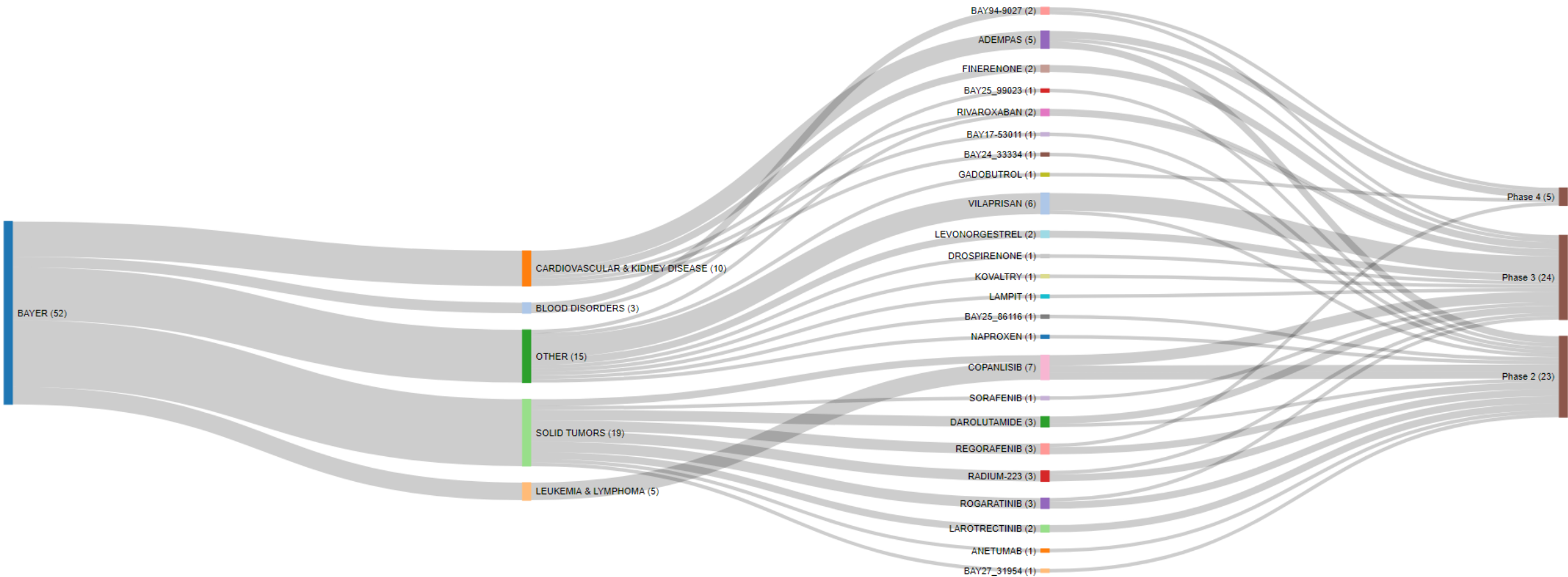
Number of on-going clinical trials for each node are displayed in parenthesis.
Clinical trials with multiple phases are counted toward the higher phase; Clinical trials with multiple clinical conditions are counted towards each condition; Clinical trials with multiple interventions are counted towards each intervention.
Note: Data analysis and data visualization performed in this paper is not an official representation of any of the sponsor's pipeline but based on the data acquired from clinicaltrials.gov.

Sankey Diagram – Amgen



Node 1: Sponsor; Node 2: Clinical Conditions; Node 3: Interventions; Node 4: Clinical Trial Phases
Number of on-going clinical trials for each node are displayed in parenthesis.
Clinical trials with multiple phases are counted toward the higher phase; Clinical trials with multiple clinical conditions are counted towards each condition; Clinical trials with multiple interventions are counted towards each intervention.
Note: Data analysis and data visualization performed in this paper is not an official representation of any of the sponsor's pipeline but based on the data acquired from clinicaltrials.gov.

Sankey Diagram – Bayer



Node 1: Sponsor; Node 2: Clinical Conditions; Node 3: Interventions; Node 4: Clinical Trial Phases
Number of on-going clinical trials for each node are displayed in parenthesis.
Clinical trials with multiple phases are counted toward the higher phase; Clinical trials with multiple clinical conditions are counted towards each condition; Clinical trials with multiple interventions are counted towards each intervention.
Note: Data analysis and data visualization performed in this paper is not an official representation of any of the sponsor's pipeline but based on the data acquired from clinicaltrials.gov.

Summary

- Using sankey diagrams for drug pipeline analysis:
 - Impressive data visualization tool to understand flow of clinical trials
 - Helps to track several clinical trials in a single view
 - Facilitates to understand weightage of clinical conditions, interventions, or phases of clinical trials
 - Represents flow in a manner that can be understood by anyone, instantly
 - Allows user to see complex pipeline of a sponsor in a single image with a focus on the clinical conditions and interventions/compounds of that sponsor
 - Makes dominant clinical conditions or interventions stand out, and they help users to see relative magnitudes and/or areas with the largest opportunities
 - Supports multiple viewing levels
- By using provided macros, sankey diagram can be adjusted as per user's need
- Users can get a high-level view, see specific details, or generate custom diagrams by using provided macros

Acknowledgement

- I would like to thank Vineet Mathur and Simon Xue for their guidance and support
- I would like to thank the dedicated people who manage and maintain clinicaltrials.gov and d3js.org

Questions?
