

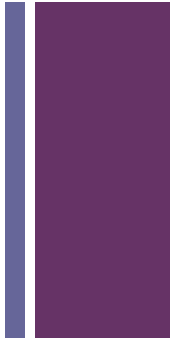


## Identifying additional samples for replication in SiGN

Brad Worrall

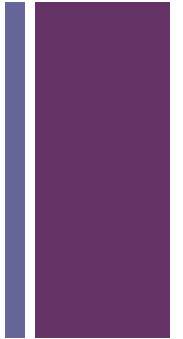
# + The power of numbers...

- The SiGN analytic committee has argued for analyzing all of the 14,000 stroke cases in SiGN in the main analyses
- Partitioning gains us little from a scientific/ analytic standpoint – lose power
- Concern that the reviewer community will “demand” replication
  - Have potential plan for replication of the all ischemic stroke phenotype
  - Challenges remain for replication of the stroke subtypes by CCS
- We know that there are other samples out there that are not currently part of SiGN





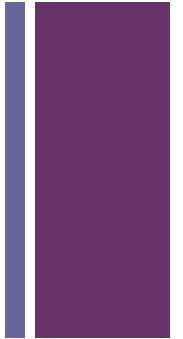
# The challenges of phenotyping...



- The grant states that CCS will be the primary subtyping system used in analyses
- Generation of CCS data was a requirement for participation and a major undertaking of the grant
- Our own data suggest that CCS and TOAST have modest correlation at best and caution against collapsing data from the two systems
- It is unlikely that there are large caches of stroke cases with CCS subtyping available that are not already identified
- Generating CCS data takes time and effort and therefore money which we do not have



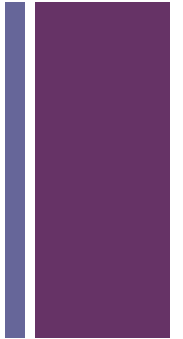
# Approaches to identifying more cases...



- Prospective recruitment - \$\$\$ and time
- Include cases from current SiGN sites that are not currently part of data set
  - Excluded cases (skewed)
  - Collected since SiGN
- Work with known sites with and without GWAS data (and a willingness to generate CCS)
  - Expansion sites – ASGC, CADISP, sifap, Finland
  - METASTROKE cases
  - COMPASS Cases
  - Other (PROFESS, SLESS, VITATOPS)
- Screen the literature for potential collaborators who have published genetic results and are not currently part of ISGC/ METASTROKE/ SiGN



# Current SiGN Sites



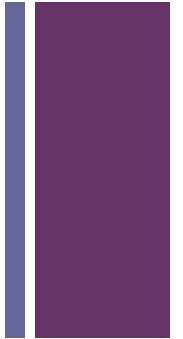
- Excluded cases
  - Some sites excluded specific subtypes from the SiGN data set (e.g. dissection or undetermined) (skewed)
    - BASICMAR has 400 cases classified as undetermined by TOAST
    - Others?
    - Skewed
  - All sites with CIDR genotyping had exclusion criteria based on CCS subtyping (undetermined, iatrogenic, rare mechanisms) or that were excluded on the bases of poor DNA or genomic data quality
    - Totals 4,264
    - Skewed



# SiGN (excluded): N = 4,264 (1,281)

Study ID	PI/contact	# cases not in SiGN	Study ID	PI/contact	# cases not in SiGN
<b>GASROS</b>	Rosand	<b>86</b>	<b>WUSTL</b>	Lee	188
<b>GCNKSS</b>	Woo	143	<b>OXVASC</b>	Rothwell	<b>90</b>
<b>GEOS</b>	Kittner/Cole	<b>5</b>	<b>Munich</b>	Dichgans	<b>323</b>
<b>ISGS</b>	Meschia	<b>116</b>	<b>Edinburgh</b>	Sudlow	<b>56</b>
<b>MCISS</b>	Grewal	246	<b>BRAINS</b>	Sharma	<b>208</b>
<b>MIAMISR</b>	Rundek	32	<b>St. George/ Cambridge</b>	Markus	<b>261</b>
<b>NHS</b>	Rexrode	154	<b>Graz</b>	Schmidt	46
<b>NOMAS</b>	Sacco	215	<b>Gothenburg</b>	Jern	285
<b>REGARDS</b>	Arnett	244	<b>Lund</b>	Lindgren	313
<b>SWISS</b>	Meschia	<b>136</b>	<b>BASICMAR</b>	Jimenez-Conde	158
<b>SPS3</b>	McLure/ Benavente	?	<b>Krakow</b>	Slowik	535
<b>WHI</b>	Smoeller	382	<b>Leuven</b>	Thijs/ Lemmons	42

# + Current SiGN Sites



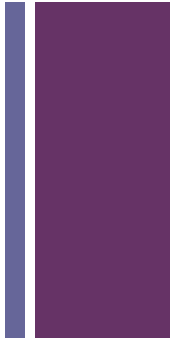
- Not previously considered for SiGN
  - Lacking genotypic data (e.g. Hugh's SLESS)
  - DNA not yet extracted
  - New recruits (e.g. MGH, Mayo)

# + SiGN (new): N = 7722 (~500 w/ CCS)

Study ID	PI/contact	# cases not in SiGN	Study ID	PI/contact	# cases not in SiGN
<b>GASROS</b>	Rosand	300	<b>WUSTL</b>	Lee	300
<b>GCKNSS</b>	Woo	500	<b>OXVASC</b>	Rothwell	500
<b>GEOS</b>	Kittner/Cole	0	<b>Munich</b>	Dichgans	
<b>ISGS</b>	Meschia	196	<b>Edinburgh</b>	Sudlow	276
<b>MCISS</b>	Grewal	100	<b>BRAINS</b>	Sharma	300
<b>MIAMISR</b>	Rundek	200	<b>St. George/ Cambridge</b>	Markus	1000
<b>NHS</b>	Rexrode	300	<b>Graz</b>	Schmidt	
<b>NOMAS</b>	Sacco	0	<b>Gothenburg</b>	Jern	300
<b>REGARDS</b>	Arnett		<b>Lund</b>	Lindgren	200
<b>SWISS</b>	Meschia	0	<b>BASICMAR</b>	Jimenez- Conde	800
<b>SPS3</b>	McLure/ Benavente	0	<b>Krakow</b>	Slowik	300
<b>WHI</b>	Smoeller	2000	<b>Leuven</b>	Thijs/ Lemmons	150



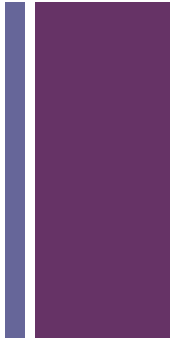
# Proposed SiGN sites with existing genotypic data



- Helsinki 2000 – 265 with genotype and TOAST
- ASGC – 900 total, 300 with CCS
- CADISP (IS) – 583 total; 583 with SS-TOAST
- Sifap – 791 total; 747 with TOAST (of 3369 IS cases)
- VISP – 2100 total; 0 with subtype
  - **Genotype N=4639**
  - **TOAST N=2495**
  - **CCS N=883 → N=1895**

# + As discussed yesterday...

- CCS data possible for these cases (infrastructure in place)
- Still takes time and money and effort
- Discussing strategies of using TOAST subtypes...





# METASTROKE: N = 14241

Study ID	PI/contact	# cases not in SiGN	Study ID	PI/contact	# cases not in SiGN
<b>ARIC</b>	Myriam Fornage	385	<b>BSS</b>	Massimo Pandolfo	225
<b>Perth</b>	Chris Levi/Graeme Hankey	400	<b>Copenhagen</b>	Børge G Nordestgaard	730
<b>Adelaide</b>	Chris Levi/Simon Koblar	300	<b>ESS</b>	Cathie Sudlow	276
<b>CHS</b>	Will Longstreth	454	<b>Glasgow</b>	Matthew Walters	675
<b>deCODE</b>	Solveig Grétarsdottir	2391	<b>Go-Darts</b>	Colin N A Palmer	737
<b>FHS</b>	Suhda Seshadri	171	<b>Interstroke</b>	Guillaume Paré	872
<b>HPS</b>	Jemma Hopewell	578	<b>Munster</b>	Andreas Gschwendtner	1232
<b>HVH</b>	Bruce Psaty	566	<b>Portugal - Oeiras</b>	Jose Ferro	534
<b>Milano</b>	Giorgio B Boncoraglio	372	<b>RACE</b>	Danish Saleheen	1322
<b>Rotterdam</b>	M Arfan Ikram	367	<b>SMART</b>	Ale Algra	632
<b>Barcelona</b>	Joan Montaner	700	<b>WHI</b>	Alex Reiner	322



# COMPASS: Cases=1156 Controls=15432



Study ID	Contact	Cases not in SiGN	Controls
ARIC	Myriam Fornage	500	2241
Health ABC	YongMei Liu	128	1100
WHI-SHARE	Cara Carty	199	8300
CHS	Josh Bis	79	641
SUGAR/SIGNET / REGARDS	Michele Sale	250	2,250

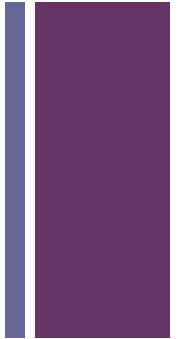
All African-American; Some have TOAST; Some not subtyped

# + Other “known” studies with DNA

## N=12632+

Study ID	PI/contact	# cases not in SiGN	Study ID	PI/contact	# cases not in SiGN
<b>Parelsnoer Project</b>	Jaap Kappelle/ Ynte Ruigrok	3000	<b>MESA</b>	Michele Sale/ Stephen Rich	120
<b>Aberdeen</b>	Mary J. MacLeod	600	<b>Dublin</b>	Peter Kelly	?
<b>Taiwan Stroke Study</b>	Jerry Rotter/ Ida Chen	3000	<b>Lille</b>	Stéphanie DeBette	?
<b>TexGen</b>	<a href="#">Fornage/ Milewicz</a>	550	<b>Lausanne</b>	Christian Wider	?
<b>VITATOPS</b>	<a href="#">Graeme Hankey</a>	800	<b>UT Houston Chronic Renal Insufficiency Cohort (CRIC)</b>	<a href="#">Fornage/ Grotta</a> Chronic Renal Insufficiency Cohort (CRIC)	?
<b>BASIC</b>	<a href="#">Lynda Lisbeth</a>	313	<b>Stroke Disparities Program</b>	Chelsea Kidwell/ Amy Hsui	?
<b>PROFESS</b>	Ralph Sacco	1200	<b>SPOTRIAS</b>	?	?
<b>Jackson Heart Study</b>	Herman Taylor	100	<b>South American Netwrok</b>	Juan Jose Martin	?
<b>Joinville Stroke Biobank</b>	Iscia Cendes	817	<b>Women's health study and WHGS</b>	Kathy Rexrode	473
<b>Physicians' Health</b>	Zee / Ridker	259	<b>Mt. Sinai</b>	Yvonne Chan	1000
<b>SLESS</b>	Hugh Markus	1000			

# + Literature review



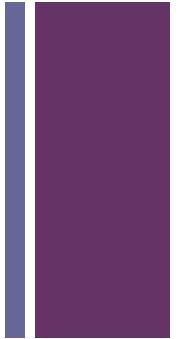
- Search terms (ischemic or ischaemic) and stroke and (genetic or genomic or genome or genetics or SNP or polymorphism) NOT (mice or mouse or animal or zebrafish or rats)
- I got just over 2200 articles and I reviewed the first ~400 articles identifying 210 articles
- Raw numbers 230,000 cases described
- 27 Meta-analyses account for 142,587
- Non-meta-analysis and non-overlapping 78,611

# + Breakdown by country

Country	Total	Unique	Meta-analysis	Country	Total	Unique	Meta-analysis
Algeria	1	1	0	Germany	9	9	0
Armenia	2	1	0	Greece	3	2	0
Australia	1	1	0	India	10	4	0
Brazil	1	1	0	Iran	1	1	0
Canada	1	1	0	Iraq	1	1	0
China	99	72	22	Israel	1	1	0
Croatia	1	1	0	Italy	4	4	0
Denmark	5	5	0	Japan	4	4	0
Finland	1	1	0	Korea	8	8	0

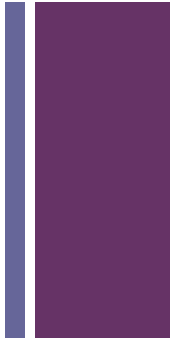


# Breakdown by country (con't)



Country	Total	Unique	Meta-analysis	Country	Total	Unique	Meta-analysi
Morocco	1	0	1	Turkey	3	3	0
Poland	4	2	0	Ukraine	1	1	0
Russia	2	2	0	USA	8	8	0
Spain	3	2	1	Zambia	1	1	0
Sweden	2	2	0				
Switzer-land	1	1	0				
Taiwan	5	5	0				
Thailand	1	1	0				
Tunisia	4	4	0				

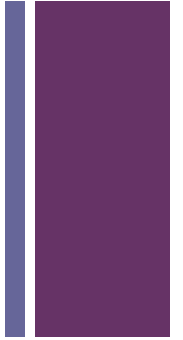
# + Next steps



- Have SiGN GRCs fill in spreadsheet regarding additional cases
- Complete initiation process for “expansion sites” – determine if in main analysis or replication
- Reach out to METASTROKE
- Reach out to COMPASS
- Reach out to other “known” groups
- Work with ISGC outreach committee to target high yield groups from literature



# Still need to resolve phenotyping



- Not an issue for the all ischemic stroke analysis
- Subtype specific analyses
  - Strategy to merge TOAST and CCS (undesirable)
  - TOAST only analyses (not c/w the grant, will lose some SiGN sites)
  - Intersection/ Union of CCS and TOAST analyses