# 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 |
|---|---|---|---|---|---|
| **GASROS** | Rosand | **86** | **WUSTL** | Lee | 188 |
| **GCNKSS** | Woo | 143 | **OXVASC** | Rothwell | **90** |
| **GEOS** | Kittner/Cole | **5** | **Munich** | Dichgans | **323** |
| **ISGS** | Meschia | **116** | **Edinburgh** | Sudlow | **56** |
| **MCISS** | Grewal | 246 | **BRAINS** | Sharma | **208** |
| **MIAMISR** | Rundek | 32 | **St. George/ Cambridge** | Markus | **261** |
| **NHS** | Rexrode | 154 | **Graz** | Schmidt | 46 |
| **NOMAS** | Sacco | 215 | **Gothenburg** | Jern | 285 |
| **REGARDS** | Arnett | 244 | **Lund** | Lindgren | 313 |
| **SWISS** | Meschia | **136** | **BASICMAR** | Jimenez-Conde | 158 |
| **SPS3** | McLure/ Benavente | ? | **Krakow** | Slowik | 535 |
| **WHI** | Smoeller | 382 | **Leuven** | 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 |
|---|---|---|---|---|---|
| **GASROS** | Rosand | 300 | **WUSTL** | Lee | 300 |
| **GCNKSS** | Woo | 500 | **OXVASC** | Rothwell | 500 |
| **GEOS** | Kittner/Cole | 0 | **Munich** | Dichgans | |
| **ISGS** | Meschia | 196 | **Edinburgh** | Sudlow | 276 |
| **MCISS** | Grewal | 100 | **BRAINS** | Sharma | 300 |
| **MIAMISR** | Rundek | 200 | **St. George/ Cambridge** | Markus | 1000 |
| **NHS** | Rexrode | 300 | **Graz** | Schmidt | |
| **NOMAS** | Sacco | 0 | **Gothenburg** | Jern | 300 |
| **REGARDS** | Arnett | | **Lund** | Lindgren | 200 |
| **SWISS** | Meschia | 0 | **BASICMAR** | Jimenez-Conde | 800 |
| **SPS3** | McLure/ Benavente | 0 | **Krakow** | Slowik | 300 |
| **WHI** | Smoeller | 2000 | **Leuven** | 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 |
|---|---|---|---|---|---|
| **ARIC** | Myriam Fornage | 385 | **BSS** | Massimo Pandolfo | 225 |
| **Perth** | Chris Levi/Graeme Hankey | 400 | **Copen-hagen** | Børge G Nordestgaard | 730 |
| **Adelaide** | Chris Levi/ Simon Koblar | 300 | **ESS** | Cathie Sudlow | 276 |
| **CHS** | Will Longstreth | 454 | **Glasgow** | Matthew Walters | 675 |
| **deCODE** | Solveig Grétarsdottir | 2391 | **Go-Darts** | Colin N A Palmer | 737 |
| **FHS** | Suhda Seshadri | 171 | **Interstroke** | Guillaume Paré | 872 |
| **HPS** | Jemma Hopewell | 578 | **Munster** | Andreas Gschwendtner | 1232 |
| **HVH** | Bruce Psaty | 566 | **Portugal - Oeiras** | Jose Ferro | 534 |
| **Milano** | Giorgio B Boncoraglio | 372 | **RACE** | Danish Saleheen | 1322 |
| **Rotterdam** | M Arfan Ikram | 367 | **SMART** | Ale Algra | 632 |
| **Barcelona** | Joan Montaner | 700 | **WHI** | 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 |
|---|---|---|---|---|---|
| **Parelsnoer Project** | Jaap Kappelle/ Ynte Ruigrok | 3000 | **MESA** | Michele Sale/ Stephen Rich | 120 |
| **Aberdeen** | Mary J. MacLeod | 600 | **Dublin** | Peter Kelly | ? |
| **Taiwan Stroke Study** | Jerry Rotter/ Ida Chen | 3000 | **Lille** | Stéphanie DeBette | ? |
| **TexGen** | Fornage/ Milewicz | 550 | **Lausanne** | Christian Wider | ? |
| **VITATOPS** | Graeme Hankey | 800 | **UT Houston** | Fornage/ Grotta | ? |
| **BASIC** | Lynda Lisbeth | 313 | **Chronic Renal Indsufficency Cohort (CRIC)** | Chronic Renal Indsufficency Cohort (CRIC) | ? |
| **PROFESS** | Ralph Sacco | 1200 | **Stroke Disparities Program** | Chelsea Kidwell/ Amy Hsui | ? |
| **Jackson Heart Study** | Herman Taylor | 100 | **SPOTRIAS** | ? | ? |
| **Joinville Stroke Biobank** | Iscia Cendes | 817 | **South American Netwrok** | Juan Jose Martin | ? |
| **Physicians' Health** | Zee / Ridker | 259 | **Women's health study and WHGS** | Kathy Rexrode | 473 |
| **SLESS** | Hugh Markus | 1000 | **Mt. Sinai** | Yvonne Chan | 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-analysi |
|---|---|---|---|---|---|---|---|
| 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