Proteomic analysis of plasma from children with sickle cell anemia and silent cerebral infarction

by Sanjay Tewari, George Renney, John Brewin, Kate Gardner, Fenella Kirkham, Baba Inusa, James E. Barrett, Stephan Menzel, Swee Lay Thein, Malcolm Ward, and David C. Rees

Proteomic analysis of plasma from children with sickle cell anemia and silent

cerebral infarction

Sanjay Tewari[1], George Renney[2], John Brewin[1], Kate Gardner[1], Fenella Kirkham[3],

Baba Inusa[4], James E Barrett[5], Stephan Menzel[1], Swee Lay Thein[6], Malcolm Ward[2]

and David C Rees[1]

1. Red Cell Biology Unit, King's College Hospital, King's College London, UK.

2. Proteomics Laboratory, Institute of Psychiatry, King's College London, UK.

3. Department of Neurosciences, Institute of Child Health, University College

   Hospital, London, UK.

4. Evelina Children's Hospital, Guy's and St Thomas' Hospital, London, UK.

5. Division of Health & Social Care Research, King's College London, UK.

6. Sickle Cell Branch, National Heart, Lung and Blood Institute, National

   Institutes of Health, Bethesda, USA.

Running head: Proteomics of silent cerebral infarction in SCA

Address for correspondence: David Rees, Department of Haematological Medicine,

King's College Hospital, Denmark Hill, London, SE5 9RS, UK.

Email: david.rees2@nhs.net.

Word count: Abstract 247, Main text 3393, 5 tables, 1 figure, 1 supplemental file.

1

Abstract

Silent cerebral infarction is the commonest neurological abnormality in children with sickle cell anemia, affecting 30-40% 14 year olds. There are no known biomarkers to identify children with silent cerebral infarcts and the pathological basis is also unknown. We used an unbiased proteomic discovery approach to identify plasma proteins differing in concentration between children with and without silent cerebral infarcts. Clinical parameters and plasma samples were analysed from 51 children (mean age 11.8 years, range 6-18) with sickle cell anemia (HbSS). 19 children had silent cerebral infarcts and 32 normal MRI; the children with silent infarcts had lower HbF levels (8.6 vs. 16.1%, P=0.049) and higher systolic blood pressures (115 vs 108.6, P=0.027). Plasma proteomic analysis showed 13 proteins increased more than 1.3 fold in the SCI patients, including proteins involved in hypercoagulability (α2-antiplasmin, fibrinogen–γ chain, thrombospondin-4), inflammation (α2-macroglobulin, complement C1s and C3) and atherosclerosis (apolipoprotein B-100). Higher levels of gelsolin and retinol-binding protein 4 were also found in the population with silent infarcts, both of which have been linked to stroke. We investigated the genetic basis of these differences by studying 359 adults with sickle cell disease (199 with silent cerebral infarcts, 160 normal MRIs), who had previously undergone a genome-wide genotyping array. None of the genes coding for the differentially expressed proteins were significantly associated with silent infarction. Our study suggests that silent cerebral infarcts in sickle cell anemia may be associated with higher systolic blood pressure, lower HbF levels, hypercoagulability, inflammation and atherosclerotic lipoproteins.

<u>Introduction</u>

Sickle cell anemia (SCA) is the commonest cause of stroke in childhood (1). The abnormal sickle hemoglobin (HbS) polymerizes when deoxygenated, damaging red cells, causing vaso-occlusion and vascular endothelial dysfunction (2). A cascade of pathological processes follows, including inflammation, hemolysis, anemia, oxidative stress, reperfusion injury, hypercoagulability, and nitric oxide deficiency (3). Survival is reduced by 20-30 years (4, 5) and vasculopathic complications include pulmonary hypertension, priapism, and cerebrovascular disease (6). The latter is the major cause of morbidity in children and takes two forms: overt stroke often associated with large vessel disease (7), and silent cerebral infarcts (SCI) of less certain pathology (8). Without intervention, overt stroke has a peak incidence of 1.02/100 patient years between 2 and 5 years (9), although primary prevention using transcranial Doppler (TCD) scanning and blood transfusion reduces this by 90% (10). The implementation of TCD-based stroke prevention has led to a significant fall in overt strokes in many countries (11, 12). SCI is more common than overt stroke with MRI showing relevant lesions in 13% of two year olds, 25% of six year olds, and 30-40% of 14 year olds(13).

Although by definition SCIs are not associated with overt symptoms, they cause significant morbidity, including a reduction in IQ, defects in executive function, epilepsy, and increased risk of further SCIs and overt strokes (13). Early detection of SCI is useful to allow cognitive assessment, educational support and therapeutic intervention. The Silent Cerebral Infarct Multi-Center Clinical Trial (SIT Trial), a randomised controlled study, showed that regular blood transfusion reduced the

incidence of recurrent infarction in children with SCIs compared to the observational arm (14).

The pathophysiology of SCI is unclear. It is often stated that SCIs are caused by small vessel disease, although there is no direct evidence for this. There are no post-mortem studies looking at the histological changes corresponding to the MRI appearances of SCI, although older studies identified small necrotic lesions in the subcortex of the brain, possibly representing SCIs (15). It is also possible that SCIs are areas of demyelination, or linked to venous sinus thrombosis (13). SCIs may also be caused by difficulties maintaining constant blood flow to the brain, leading to watershed infarction precipitated by acute anemia or hypoxia(16). The lack of pathophysiological understanding makes it difficult to develop new therapeutic strategies. Here we used an unbiased proteomic approach to identify plasma proteins associated with SCI and potentially linked to the underlying pathophysiology.

It is not easy to identify children with or at increased risk of SCI. Baseline data from the SIT Trial showed that steady state hemoglobin <7.6g/dl, steady state systolic blood pressure >104mmHg and male sex were all associated with increased risk of SCI, although a model combining all three significant factors had weak predictive powers, with a C statistic of about 0.6 (8). Abnormal transcranial Doppler velocities are not reliably associated with SCI, although there is a possible association with stenosis of the cervical internal carotid artery (17). Neuropsychometry can detect some children with SCIs, but is time consuming and not easily available. MRI is

currently the only way of diagnosing SCI, although before 7 years of age this usually requires sedation or general anesthetic, which carries significant risks in children with SCD and is unsuitable for screening. MRI is also not widely available in many African countries where SCA is most prevalent. It would therefore be beneficial to develop a simple, screening test for SCI, potentially based on blood testing.

Proteomics can detect, identify and quantify very low concentrations of proteins. For example, a recent study suggested that urine proteomics might help in the diagnosis of acute stroke by detecting brain peptides released by ischemia (18). The presence of SCI may suggest small blood vessel disease and a tendency to infarct brain tissue, which is likely to be a chronic process, releasing detectable brain proteins into plasma. Plasma protein profiles could also differ between those with and without SCI because of causative pathological factors, including coagulopathy, inflammation, hypoxia and vasculopathy; these differences may be acquired or inherited, and reflect genetic predisposition to silent cerebral infarction. We investigated the possibility that there might be differences in the plasma proteome between children with and without SCIs, with a view to understanding more about the pathophysiology of SCIs and identifying potentially useful biomarkers. We further investigated the hypothesis that genetic factors account for the differences in concentrations of the various proteins, using data from a genome wide association study (GWAS) in a separate cohort of patients with SCA and SCIs.

Methods

Patients and setting

The National Research Ethics Committee approved the study (reference

13/LO/0709). Children were recruited from clinics at King's College Hospital and the

Evelina Children's Hospital. Approximately 1000 different children are seen annually

in these clinics with a further 500 seen in outreach clinics (19). The aim was to recruit

a total of 50 patients, with equal numbers of those with SCIs and controls with

normal brain MRIs. Children known to have SCIs from previous MRI scans were

selectively recruited. Eligible patients without previous MRIs were recruited

sequentially in clinic, on the basis that 20-30% of these children would have SCIs, and

combining the two groups would result in approximately 25 children in each group.

SCI and control groups were not specificaly matched for age and gender. Inclusion

criteria were: sickle cell anemia (HbSS), age 8-18 years old, normal or conditional

TCD velocities (<200cm/s) (10), able to tolerate MRI scan without sedation. Exclusion

criteria were: history of overt stroke, blood transfusion in last 4 months, serious co-

existing disorder (renal failure, HIV, hepatitis, malignancy, autoimmune disease,

chronic infection, cardiac disease, long-term medication), acute complications

requiring hospital attendance in last 3 months. Patients on hydroxyurea (HU) were

recruited to both arms. Routine clinical and laboratory data were collected.

Brain MRI and neurological assessment

Brain MRIs were performed to identify the presence of SCIs using standardised

definitions (20) (see Supplementary Materials). On the day of blood sampling, all

children underwent standardised neurological examination by experienced clinicians

to document the absence of neurological abnormalities suggestive of an overt stroke. All children had routine TCD scans performed at least annually, and these were used in this study.

Proteomic Analysis

Blood samples were collected on all children at the time of consent. A data-driven proteomics approach was used to discover protein changes in the plasma associated with SCI. We used an established workflow combining isobaric Tandem Mass Tags (TMT) reagents with off gel fractionation, followed by high sensitivity rapid throughput mass spectrometry (MS) and LC-MS/MS. (details in Supplementary Materials). Relative amounts of each peptide were compared between the SCI and control groups, and P values were calculated for those peptides showing greater than 1.3 fold differences.

Targeted plasma analysis for biomarkers of neurological disease

Plasma samples were also analysed for 92 protein biomarkers known to be implicated in neurological disease, using a multiplexed proximity extension assay(21) (see Supplementary Materials for full list of biomarkers).

Data Analysis

Statistical significance was assessed using appropriate tests to compare means with adjustments for multiple comparisons when appropriate. Statistical analyses were performed in the proteomics laboratory and medical statistics department of Public Health Sciences at King's College London.

Genetic association of protein biomarkers with SCIs

The genes coding for plasma proteins which differed significantly between children with and without SCIs were identified using standard databases. Genetic association analysis was then undertaken in a separate adult cohort at KCH, evaluating the variants within each gene and their association with SCI (22) (details in Supplementary Materials).

Results

55 patients consented to take part in the study, although no blood samples were taken from one, and three further patients did not have an MRI scan performed. Plasma from 54 patients underwent proteomic analysis, although only data from the 51 patients (22 female) with MRI results were analysed. 19 (37%) had SCIs and 32 (63%) normal brain MRIs, deviating slightly from the aim of equal numbers in each groups because of recruitment of children without previous MRIs, the majority of whom had normal brain imaging (Figure 1). 15 children were identified as having SCIs from scans performed as part of this study, and four children were recruited as a result previous MRI scans showing SCIs; in these four children, there was a median delay of 20 months (range 9 – 40 months) between MRI scan and blood sampling for proteomic analysis. One child had conditional velocities and 50 normal velocities on TCD scanning.

Basic clinical and laboratory data

Relatively more males (13/29, 44%) than females (6/22, 27%) were in the SCI group, although this was not significantly different (Chi square, P=0.199). There was no significant difference in HU use between the SCI and control groups: controls 11/32 (36%) on HU, SCI 7/19 (37%) on HU (Chi square, P = 0.86). G6PD assays were available for 45 patients, and there was no association with SCI: G6PD deficiency was present in 3/29 (10%) controls and 2/16 (12%) with SCIs (Chi square, P=0.83). Basic laboratory and clinical characteristics are summarized in Table 1. Mean HbF levels were significantly lower (6.1% versus 8.6%, P=0.049), and mean systolic blood pressures significantly higher (115 versus 108.6, P=0.027) in those with SCIs compared to controls.

Plasma proteomic analysis

A total of 4662 different peptides were identified, although only 1312 were present in all 51 patient samples, corresponding to 346 different proteins. 44 peptides from 13 different proteins were present at more than 1.3 times the concentration in the SCI patients compared to controls (Table 2), whereas 41 peptides from 4 proteins were more than 1.3 higher in the control population (Table 3). Proteins differentially expressed in the SCI group included those implicated in coagulation, lipid and inflammatory pathways.

Genetic association of identified biomarkers with SCI

We explored the possibility that differences in concentrations of proteins (Tables 2 and 3) between SCI and control populations were inherited, using a different sample of adult patients who had undergone brain MRI and a genome wide genotyping

9

array. Of 359 patients with brain MRIs and MEGAchip data, 199 had SCI and 160 did not. Regions of interest were analysed for the candidate genes identified in the proteomic study; the most significant genetic variant from each region is shown in Table 4. No genetic variant in the candidate genes was significantly associated with SCI after correcting for multiple testing. A variant in gelsolin was closest to achieving significance (P=0.00029, threshold for significance 0.00009).

Targeted plasma analysis

Concentrations of 92 known neurological biomarkers were measured in the plasma of 54 patients who consented to the study and gave blood samples. 51 patients had MRI scans available and were analysed. In order to control for the potential effect of hydroxyurea, a logistic regression model was fitted for each biomarker; the dependent variable was SCI status, and the biomarker and hydroxyurea were included as covariates in each model. Seven biomarkers were significantly associated with SCI, but after correction for multiple hypotheses none of these were significant (Table 5). Several biomarkers were strongly correlated with each other and a heatmap of the correlation matrix indicated a moderate level of redundancy between them. A principal component analysis found that approximately 90% of the variance could be attributed to the first 30 principal components. None of the principal components were found to be differentially expressed with respect to SCI status, either with or without controlling for HU status.

Discussion

Our study found that the presence of SCI in SCA is associated with higher systolic blood pressures, an association that was also found in the SIT trial (8). Blood pressure is relatively low in SCA (23), possibly related to chronic anemia and subsequent vasodilatation, and the levels found in our study, with a mean of 115mmHg in the SCI group, would not be considered hypertensive. The exact way in which relative systolic hypertension is associated with SCI in children with SCA is unknown, although systolic hypertension is linked with other vascular complications in SCA (overt stroke (23), renal insufficiency (24), and high tricuspid jet velocity (24)). Systolic hypertension is also a risk factor for SCI in the non-sickle population (25). The association between relative systolic hypertension and SCI in SCA seems robust, and is biologically plausible, although there is currently no evidence on whether it is beneficial to lower blood pressure in children with SCA, and at what level such treatment should be considered.

We also found that higher HbF levels were protective against SCI. Increased HbF levels lessen the severity of SCA, with a clear association with decreased pain and longer life expectancy (2), but a more complex association with cerebrovascular complications. In the SIT Trial, HbF was not associated with SCI in a logistic regression model (8) and analysis of data from the Cooperative Study of Sickle Cell Disease also showed no protective effect of higher HbF levels (26). Other studies have suggested that there may be neurological benefit to having higher HbF levels. In France, HbF levels were higher in children without SCIs compared to those with (16.5% vs. 13.1%, P=0.02) (27); a Dutch study found that although HbF did not protect against the occurrence of SCIs, higher levels were associated with a smaller

volume of white matter hyperintensities (28); a study on Brazilian patients showed protection against stroke was associated with genetic determinants of high HbF levels, particularly *BCL11A* alleles (29). It seems likely that the inherited ability to make more HbF offers some protection against neurological injury in SCA, although this may not always be reflected in the measured HbF percentage, particularly in small studies.

The unbiased proteomic discovery part of this study identified 44 peptides from 17 different proteins which were significantly up or down regulated compared to levels in the control population, without correction for multiplicity (Tables 2 and 3). Many of these proteins can be plausibly implicated in pathogenesis of SCI and fall in to three broad groups.

Prothrombotic proteins

Four proteins which were found at higher levels in the SCI patients are known to be prothrombotic. α_2-antiplasmin is a serine protease inhibitor which inactivates plasmin and decreases the rate of fibrinolysis; high levels are associated with increased ischemic stroke in the non-sickle population (30). Similarly, α_2-macroglobulin has prothrombotic properties via the inactivation of plasmin, and elevated levels have been linked to stroke and white matter lesions in non-sickle patients(31). The fibrinogen γ chain was increased 1.33 in the SCI population, and is at the center of fibrin polymerization, interaction with platelets and regulation of factor XIII activity (32). Increased levels have been associated with vasculopathy, higher levels being increased in inflammation and also possibly inherited (33).

Fourthly, thrombospondin-4 was increased 1.32 fold with SCI. Thrombospondins mediate cellular adhesion to other cells and matrix, and are implicated in many thrombotic and vasculopathic processes. Thrombospondin-1 levels have been found to be increased in children with SCA in association with both SCI(34) and overt stroke (35), although thrombospondin-4 has not previously been studied in this context. These findings support the potential use of anticoagulant and antiplatelet agents to prevent and treat SCI in SCA.

Pro-inflammatory proteins

Elevated levels of α_2-macroglobulin, complement C1s and complement C3 in the SCI patients suggest that there is increased inflammation, which is known to predispose towards vasculopathy and stroke. As mentioned previously, α_2-macroglobulin is a prothrombotic protease inhibitor, but is also involved in inflammatory states, as a carrier for interleukin-6(31). Both complements C1s and C3 are elevated in inflammation, and high levels of C3 are a known risk factor for atherosclerosis and stroke in the general population(36). Inflammation is known to be an important component of pathophysiology in SCA, and the role of anti-inflammatory agents in preventing SCIs should be explored further.

Lipoproteins

Apolipoprotein B-100 is the main component of low density lipoprotein cholesterol (LDL-C) and high levels are a well-established risk factor for atherosclerosis. Various studies have shown that drugs which reduce LDL-C also reduce the rate of cardiovascular events, including ischemic stroke (37), suggesting a possible

therapeutic option for children with SCA and ischemic stroke for drugs such as statins. We also found that apolipoprotein A-IV levels were higher in the SCI population, which is perhaps paradoxical in that it is thought to be protective against coronary syndromes (38), although it has not been directly implicated in cerebrovascular disease or studied in children. A recent study suggested that higher levels were associated with reduced glomerular filtration rates (39), and this could be one explanation for the differences in our study.

Other interesting candidates

Several other interesting candidates were identified, both shedding light on the pathophysiology of SCI and with the potential to act as useful biomarkers. Gelsolin binds to and remodels actin by both severing and capping the protein; it is present in both cytoplasm and plasma. Increased plasma levels have been identified as a poor prognostic marker in ischemic stroke (40). Further significant actions in the context of SCI include a role in promoting platelet formation and activation (41). In our genetic analysis, a genetic variant in the gelsolin gene showed a trend towards being associated with SCI.

Retinol binding protein-4 is a lipocalin transporting vitamin A from the liver to other tissues, including the brain; plasma levels have been found to be elevated in atherosclerosis and also ischemic stroke (42, 43).

Pigment epithelium-derived factor is multifunctional with anti-angiogenic properties, secreted by different tissues including adipocytes. It is potentially significant in the

development of SCI because of its ability to induce insulin resistance, and inflammation and proliferation of muscle cells (44). It is harder to know the significance of some other proteins, including the lower levels of immunoglobulin chains and platelet basic protein in SCI, although immunodeficiency has been associated with primary and recurrent stroke in the general pediatric population(45).

The genetic association analysis did not suggest that any of the protein differences seen had a strong genetic basis, although the pathophysiology of SCIs seen in the adults used in this part of the study may differ from that in the children used in the proteomic part.

Several plausible biomarkers for SCI were identified in the targeted plasma analysis, including von Willebrand factor C domain, although none of these reached statistical significance when corrected for multiple assays. Similarly, none of the candidate proteins identified by proteomics showed a convincing genetic association with SCI in a separate cohort of 359 patients with SCA, although the pathophysiology of SCIs seen in the adults used in this part of the study may differ from that in the children used in the proteomic part. This may also reflect the relatively small size of this study, or that the changes seen in proteins of those with SCIs are secondary to other pathological events rather than inherited quantitative trait loci.

The strength of our study is that it used an unbiased proteomics approach to identify proteins involved in the pathogenesis of SCI, together with a combination of targeted plasma analysis and genetic association studies. Our findings suggest that

15

proinflammatory and prothrombotic states contribute to the development of SCI, with possible abnormalities in lipoproteins. Previous studies to identify biomarkers for SCI in SCD have been summarized by Lance et al (35). They identified nine studies measuring a range of different candidates, although no studies used an unbiased proteomic approach. 17 potential candidates were identified in this review, again with a predominance of factors suggesting coagulation activation and increased inflammation (35). None of the identified candidates were the same as our candidates, although one study found increased levels of thrombospondin-1 in children with SCIs and SCA(34).

Although our study is relatively small, in keeping with proteomic discovery approaches, some significant factors have emerged. We confirmed the importance of hypertension as a risk factor for SCI, together with support for higher HbF levels being protective, suggesting that anti-hypertensives and hydroxyurea may be of benefit. The utility of individual biomarkers, such as gelsolin or α-2-antiplasmin, to act as clinically useful biomarkers in screening for SCI needs to be confirmed in larger, prospective studies. Previous studies have shown increased inflammation and coagulation in children with SCD(46, 47), possibly linked to small blood vessel disease(48), and our study suggests that this is more marked in children who also have SCIs. This relatively prothrombotic and pro-inflammatory state associated with SCIs suggests that it may be useful to study anti-inflammatory and antiplatelet agents in clinical trials of SCI prevention. On-going clinical trials of antiplatelet agents(49), statins(50) and canakinumab are particularly relevant in this context.

References

1.   Earley CJ, Kittner SJ, Feeser BR, et al. Stroke in children and sickle-cell disease: Baltimore-Washington Cooperative Young Stroke Study. Neurology. 1998;51(1):169-176.

2.   Brousse V, Makani J, Rees DC. Management of sickle cell disease in the community. BMJ. 2014;348:g1765.

3.   Rees DC, Gibson JS. Biomarkers in sickle cell disease. Br J Haematol. 2012;156(4):433-445.

4.   Platt OS, Brambilla DJ, Rosse WF, et al. Mortality in sickle cell disease. Life expectancy and risk factors for early death. N Engl J Med. 1994;330(23):1639-1644.

5.   Gardner K, Douiri A, Drasar E, et al. Survival in adults with sickle cell disease in a high-income setting. Blood. 2016;128(10):1436-1438.

6.   Kato GJ, Gladwin MT, Steinberg MH. Deconstructing sickle cell disease: reappraisal of the role of hemolysis in the development of clinical subphenotypes. Blood Rev. 2007;21(1):37-47.

7.   Helton KJ, Adams RJ, Kesler KL, et al. Magnetic resonance imaging/angiography and transcranial Doppler velocities in sickle cell anemia: results from the SWiTCH trial. Blood. 2014;124(6):891-898.

8.   DeBaun MR, Sarnaik SA, Rodeghier MJ, et al. Associated risk factors for silent cerebral infarcts in sickle cell anemia: low baseline hemoglobin, sex, and relative high systolic blood pressure. Blood. 2012;119(16):3684-3690.

9.   Ohene-Frempong K, Weiner SJ, Sleeper LA, et al. Cerebrovascular accidents in sickle cell disease: rates and risk factors. Blood. 1998;91(1):288-294.

10.   Adams RJ, McKie VC, Hsu L, et al. Prevention of a first stroke by transfusions in children with sickle cell anemia and abnormal results on transcranial Doppler ultrasonography. N Engl J Med. 1998;339(1):5-11.

11.   Bernaudin F, Verlhac S, Arnaud C, et al. Impact of early transcranial Doppler screening and intensive therapy on cerebral vasculopathy outcome in a newborn sickle cell anemia cohort. Blood. 2011;117(4):1130-1140; quiz 1436.

12.   Fullerton HJ, Adams RJ, Zhao S, Johnston SC. Declining stroke rates in Californian children with sickle cell disease. Blood. 2004;104(2):336-339.

13.   DeBaun MR, Armstrong FD, McKinstry RC, Ware RE, Vichinsky E, Kirkham FJ. Silent cerebral infarcts: a review on a prevalent and progressive cause of neurologic injury in sickle cell anemia. Blood. 2012;119(20):4587-4596.

14.   DeBaun MR, Gordon M, McKinstry RC, et al. Controlled trial of transfusions for silent cerebral infarcts in sickle cell anemia. N Engl J Med. 2014;371(8):699-710.

15.   Rothman SM, Fulling KH, Nelson JS. Sickle cell anemia and central nervous system infarction: a neuropathological study. Ann Neurol. 1986;20(6):684-690.

16.   Dowling MM, Quinn CT, Plumb P, et al. Acute silent cerebral ischemia and infarction during acute anemia in children with and without sickle cell disease. Blood. 2012;120(19):3891-3897.

17.   Deane CR, Goss D, Bartram J, et al. Extracranial internal carotid arterial disease in children with sickle cell anemia. Haematologica. 2010;95(8):1287-1292.

18.   Dawson J, Walters M, Delles C, Mischak H, Mullen W. Urinary proteomics to support diagnosis of stroke. PLoS One. 2012;7(5):e35879.

19.   Day TG, Thein SL, Drasar E, et al. Changing pattern of hospital admissions of children with sickle cell disease over the last 50 years. J Pediatr Haematol Oncol. 2011;33(7):491-495.

20.   Casella JF, King AA, Barton B, et al. Design of the silent cerebral infarct transfusion (SIT) trial. Pediatr Hematol Oncol. 2010;27(2):69-89.

21.   Assarsson E, Lundberg M, Holmquist G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS One. 2014;9(4):e95192.

22.   Cheverud JM, Vaughn TT, Pletscher LS, et al. Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. Mamm Genome. 2001;12(1):3-12.

23.   Rodgers GP, Walker EC, Podgor MJ. Is "relative" hypertension a risk factor for vaso-occlusive complications in sickle cell disease? Am J Med Sci. 1993;305(3):150-156.

24.   Gordeuk VR, Sachdev V, Taylor JG, Gladwin MT, Kato G, Castro OL. Relative systemic hypertension in patients with sickle cell disease is associated with risk of pulmonary hypertension and renal insufficiency. Am J Hematol. 2008;83(1):15-18.

25.   Vermeer SE, Koudstaal PJ, Oudkerk M, Hofman A, Breteler MM. Prevalence and risk factors of silent brain infarcts in the population-based Rotterdam Scan Study. Stroke. 2002;33(1):21-25.

26.   Kinney TR, Sleeper LA, Wang WC, et al. Silent cerebral infarcts in sickle cell anemia: a risk factor analysis. The Cooperative Study of Sickle Cell Disease. Pediatrics. 1999;103(3):640-645.

27.   Bernaudin F, Verlhac S, Arnaud C, et al. Chronic and acute anemia and extracranial internal carotid stenosis are risk factors for silent cerebral infarcts in sickle cell anemia. Blood. 2015;125(10):1653-1661.

28.   van der Land V, Mutsaerts HJ, Engelen M, et al. Risk factor analysis of cerebral white matter hyperintensities in children with sickle cell disease. Br J Haematol. 2016;172(2):274-284.

29.   Leonardo FC, Brugnerotto AF, Domingos IF, et al. Reduced rate of sickle-related complications in Brazilian patients carrying HbF-promoting alleles at the BCL11A and HMIP-2 loci. Br J Haematol. 2016;173(3):456-460.

30.   Reed GL, Houng AK, Singh S, Wang D. alpha2-Antiplasmin: New Insights and Opportunities for Ischemic Stroke. Semin Thromb Hemost. 2017;43(2):191-199.

31.   Nezu T, Hosomi N, Aoki S, et al. Alpha2-macroglobulin as a promising biomarker for cerebral small vessel disease in acute ischemic stroke patients. J Neurol. 2013;260(10):2642-2649.

32.   Mosesson MW. Fibrinogen gamma chain functions. J Thromb Haemost. 2003;1(2):231-238.

33.   Williams SR, Hsu FC, Keene KL, et al. Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. Neurology. 2016;86(4):351-359.

34.   Faulcon LM, Fu Z, Dulloor P, et al. Thrombospondin-1 and L-selectin are associated with silent cerebral infarct in children with sickle cell anaemia. Br J Haematol. 2013;162(3):421-424.

35.   Lance EI, Casella JF, Everett AD, Barron-Casella E. Proteomic and biomarker studies and neurological complications of pediatric sickle cell disease. Proteomics Clin Appl. 2014;8(11-12):813-827.

36.   Niculescu F, Rus H. The role of complement activation in atherosclerosis. Immunol Res. 2004;30(1):73-80.

37.   Ray KK, Ginsberg HN, Davidson MH, et al. Reductions in Atherogenic Lipids and Major Cardiovascular Events: A Pooled Analysis of 10 ODYSSEY Trials Comparing Alirocumab With Control. Circulation. 2016;134(24):1931-1943.

38.   Kronenberg F, Stuhlinger M, Trenkwalder E, et al. Low apolipoprotein A-IV plasma concentrations in men with coronary artery disease. J Am Coll Cardiol. 2000;36(3):751-757.

39.   Mack S, Coassin S, Vaucher J, Kronenberg F, Lamina C, Apo AIVGC. Evaluating the Causal Relation of ApoA-IV with Disease-Related Traits - A Bidirectional Two-sample Mendelian Randomization Study. Sci Rep. 2017;7(1):8734.

40.   Garcia-Berrocoso T, Penalba A, Boada C, et al. From brain to blood: New biomarkers for ischemic stroke prognosis. J Proteomics. 2013;94:138-148.

41.   Silacci P, Mazzolai L, Gauci C, Stergiopulos N, Yin HL, Hayoz D. Gelsolin superfamily proteins: key regulators of cellular functions. Cell Mol Life Sci. 2004;61(19-20):2614-2623.

42.   Llombart V, Garcia-Berrocoso T, Bustamante A, et al. Plasmatic retinol-binding protein 4 and glial fibrillary acidic protein as biomarkers to differentiate ischemic stroke and intracerebral hemorrhage. J Neurochem. 2016;136(2):416-424.

43.   Sasaki M, Otani T, Kawakami M, Ishikawa SE. Elevation of plasma retinol-binding protein 4 and reduction of plasma adiponectin in subjects with cerebral infarction. Metabolism. 2010;59(4):527-532.

44.   Famulla S, Lamers D, Hartwig S, et al. Pigment epithelium-derived factor (PEDF) is one of the most abundant proteins secreted by human adipocytes and induces insulin resistance and inflammatory signaling in muscle and fat cells. Int J Obes (Lond). 2011;35(6):762-772.

45.   Ganesan V, Prengler M, Wade A, Kirkham FJ. Clinical and radiological recurrence after childhood arterial ischemic stroke. Circulation. 2006;114(20):2170-2177.

46.   van der Land V, Peters M, Biemond BJ, Heijboer H, Harteveld CL, Fijnvandraat K. Markers of endothelial dysfunction differ between subphenotypes in children with sickle cell disease. Thromb Res. 2013;132(6):712-717.

47.   Chen J, Hobbs WE, Le J, Lenting PJ, de Groot PG, Lopez JA. The rate of hemolysis in sickle cell disease correlates with the quantity of active von Willebrand factor in the plasma. Blood. 2011;117(13):3680-3683.

48.   Colombatti R, De Bon E, Bertomoro A, et al. Coagulation activation in children with sickle cell disease is associated with cerebral small vessel vasculopathy. PLoS One. 2013;8(10):e78801.

49.   Heeney MM, Hoppe CC, Abboud MR, et al. A Multinational Trial of Prasugrel for Sickle Cell Vaso-Occlusive Events. N Engl J Med. 2016;374(7):625-635.

50.   Hoppe C, Jacob E, Styles L, Kuypers F, Larkin S, Vichinsky E. Simvastatin reduces vaso-occlusive pain in sickle cell anaemia: a pilot efficacy trial. Br J Haematol. 2017;177(4):620-629.

	Controls			Silent cerebral infarction			P value
	n	mean	s.d.	n	mean	s.d.	
age (years)	32	11.5	3.17	19	12.4	3.82	0.40
Hb (g/l)	32	83.4	8.45	19	86.1	11.1	0.33
HbF (%)	**32**	**8.6**	**5.1**	**18**	**6.1**	**3.6**	**0.049**
MCH (pg)	32	28.0	4.7	19	28.6	4.0	0.68
neutrophils (x10^9/l)	32	4.7	1.6	19	4.4	1.8	0.55
reticulocytes (%)	27	12.3	3.7	15	12.4	4.7	0.99
bilirubin	32	40.5	15.5	19	42.9	20.5	0.64
AST	32	57.5	26.0	19	56.0	19.0	0.84
Creatinine	32	31.9	9.9	19	36.4	11.8	0.14
LDH	32	567	135	18	565	148	0.97
Diastolic bp (mmHg)	26	66.5	6.9	14	66.5	4.9	0.99
Systolic bp (mmHg)	**26**	**108.6**	**7.6**	**14**	**115**	**9.6**	**0.027**

Table 1: Comparison of laboratory and clinical measurements of those with and without SCI. P values from t-tests, with significant results shown in bold (P<0.05), not corrected for multiple comparisons.

Protein	Number different peptides	Normalized concentration in controls	Normalized concentration in SCI	Fold change in SCI	P value
α-2-antiplasmin	3	0.635	0.914	1.44	0.004
Albumin	1	0.505	0.703	1.39	0.005
α-2-macroglobulin	7	0.601	0.832	1.38	0.01
Apolipoprotein A-IV	7	0.28	0.421	1.50	0.0002
Apolipoprotein B-100	1	0.303	0.428	1.41	0.014
Complement C1s	1	0.42	0.565	1.34	0.014
Complement C3	2	0.414	0.561	1.35	0.004
Fibrinogen γ chain	1	0.936	1.25	1.33	0.01
Gelsolin	10	0.411	0.577	1.40	0.00006
Inter-α–trypsin inhibitor heavy chain-1	2	0.833	1.14	1.36	0.002
Pigment epithelium-derived factor	1	1.14	1.5	1.32	0.006
Retinol-binding protein 4	3	0.465	0.625	1.34	0.002
Thrombospondin-4	2	0.577	0.763	1.32	0.001

Table 2: List of all peptides and proteins present at significantly increased concentrations (>1.3 fold) in children with SCIs compared to controls. Where more than one peptide was identified, the most significant difference is given. Concentrations are normalized against standard controls.

Protein	Number different peptides	Normalized concentration in controls	Normalized concentration in SCI	Fold change in SCI	P value
Ig alpha-1 chain	1	0.757	0.541	0.71	0.021
Ig gamma-1 chain	35	1.74	1.09	0.63	0.018
Ig gamma-3 chain	1	1.08	0.821	0.76	0.035
Platelet basic protein	4	1.865	1.27	0.68	0.015

Table 3: List of all peptides and proteins present at significantly decreased

concentrations (>0.77 fold) in children with SCIs compared to controls. Where more

than one peptide was identified, the most significant difference is given.

Concentrations are normalized against standard controls.

| Protein	Gene Symbol	Chromosome	Number of SNPs analysed	P value of most significant genetic variant
alpha-2-macroglobulin	A2M	9	431	0.0140
albumin	ALB	4	89	0.0016
apoliprotein B-100	APOB	2	180	0.0587
complement C1s	C1S	12	360	0.0016
complement C3	C3	19	449	0.0257
Fibrinogen gamma chain	FGG	4	23	0.0258
gelsolin	GSN	9	607	0.0003
Ig alpha-1 chain	IGHA1	14	25	0.2699
Ig gamma-1 chain	IGHG1	14	43	0.0609
Ig gamma-3 chain	IGHG3	14	26	0.0488
inter-alpha-trypsin inhibitor heavy chain	ITIH1	3	75	0.0180
pigment epithelium-derived factor	PEDF/SERPINF1	17	117	0.0495
platelet basic protein	PPBP	4	4	0.4393
retinol-binding protein 4	RBP4	10	99	0.0040
alpha-2-antiplasmin	SERPINF2	17	70	0.0233
thrombospondin 4	THBS4	5	387	0.0207

Table 4: Association of candidate genes for SCI identified in the proteomic part of the study. The most significant genetic variants are shown. None of the variants were significant after adjustment for multiple comparisons.

	Unadjusted P value	Corrected P value
Von Willebrand factor C domain containing Protein 2	0.03	0.62
Smooth muscle-associated protein 2	0.02	0.62
ADAM metalloproteinase domain 22	0.04	0.62
ADAM metalloproteinase domain 23	0.04	0.62
Scavenger receptor class A member 5	0.03	0.62
TNF receptor superfamily member 12A	0.02	0.62
Bone morphogenetic protein 4	0.03	0.62

Table 5: Known biomarkers of neurological disease showing significant differences between SCI and control populations, although none of these maintain significance when correcting for multiple tests.

Figure Legend

Figure 1: Diagram showing the flow of patients through the study.
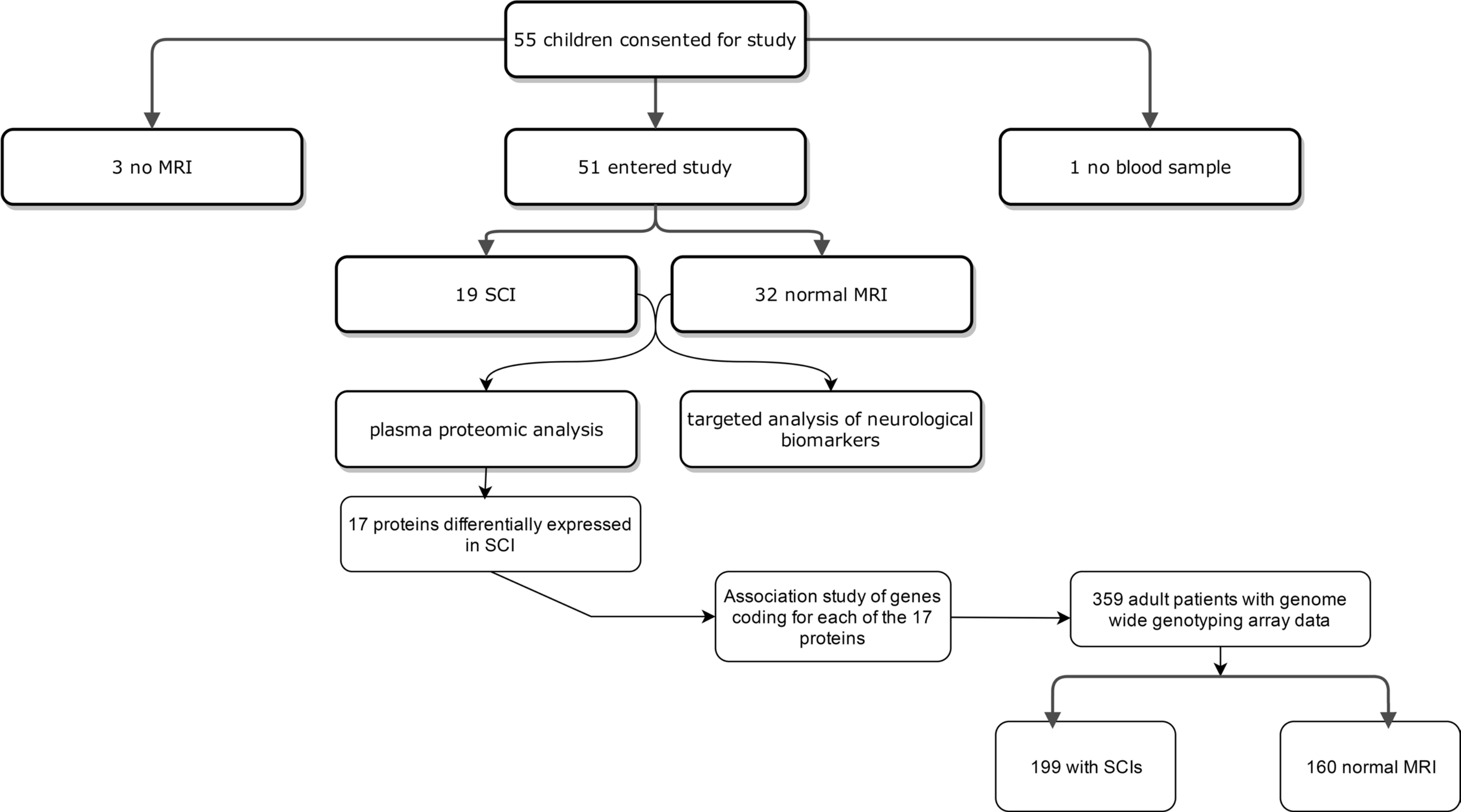
Brain MRI and Neurological Examination

Many of the eligible children for this study had previously had brain MRI for routine clinical or research purposes. MRI results were considered eligible for entry into this study if performed in the 12 months preceding the date of consent. Children known to have SCIs based on previous MRIs were selectively recruited to the study. Children without an MRI in the last 12 months had an MRI scan performed specifically as part of this study within 4 weeks of blood sampling, and were assigned to either the control or SCI group. A pediatric neuroradiologist with experience of interpreting MRIs in children with sickle cell disease reported all scans. SCI was defined as a signal abnormality visible on axial and coronal planes on the T2-weighted images, which is at least 3mm in one dimension, without a history of overt stroke or a neurological deficit corresponding to the MRI lesion.

Proteomic Methods

TMT®s (Proteome Sciences plc) are a set of isobaric mass tags for labeling proteins and peptides at amine functions and subsequent mixing of up to ten different protein samples. During MS/MS, each set of tags gives rise to different reporter ions with specific molecular masses thus allowing them to share the same chemical structure but have different molecular weights dependent on the different number of heavy isotopes contained. Peptides labelled with different TMTs show identical chromatographic retention times as well as identical ionization and fragmentation behavior of tags and tagged species.

The proteins were initially digested into peptides using trypsin, and TMT labeling then performed. Off gel fractionation was then used to separate the peptides prior to LC-MS/MS. Protein discovery study was performed in a reference-controlled manner and this approach generated considerable coverageof the plasma proteome(45).

Mass spectral data were acquired using an Orbitrap Velos mass spectrometer (Thermo Scientific). Compilation and validation of the peptide and protein assignments were carried out using Proteome Discoverer v1.3 (Thermo) and Scaffold v2.6 (Proteome Software Inc) mass spectrometry software. Proteins were quantified based on the intensities of the reporter ions released from the TMT tags during MS/MS of peptides. The use of the TMT10plex reagents enabled the simultaneous analysis of five samples alongside an overall study reference comprising of a pool of all subjects. The proteomics study comprised 5 separate TMT10plex experiments, each with 10 samples and a reference channel for normalization purposes.

Data were exported for each TMT set into an Excel file, and duplicate peptides were removed. The data were filtered so that only peptides identified in all 10 sample channels were analysed further. The data were normalized and merged against the peptide sequence using R software. The merged data were further filtered such that only peptides present in all patient samples were studied. Relative amounts of each peptide were compared between the SCI and control groups, and P values calculated, for those samples with >1.3 fold differences.

Neurological biomarkers included multiplexed proximity extension assay

Plasma samples, anticoagulated with EDTA, were frozen at -80°C within 6 hours of venesection, and transported as one batch to Olink Proteomics in Sweden (www.olink.com/proseek-multiplex) on dry ice for analysis using their Proseek Multiplex Neurology I panel, which includes the following markers:

ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 (CD38)

Alpha-2-macroglobulin receptor-associated protein (Alpha-2-MRAP)

BDNF/NT-3 growth factors receptor (NTRK2)

Beta-nerve growth factor (Beta-NGF)

Bone morphogenetic protein 4 (BMP-4)

Brain-derived neurotrophic factor (BDNF)

Brevican core protein (BCAN)

Brorin (VWC2)

C-type lectin domain family 1 member B (CLEC1B)

C-type lectin domain family 10 member A (CLEC10A)

Cadherin-3 (CDH3)

Cadherin-6 (CDH6)

Carboxypeptidase A2 (CPA2)

Carboxypeptidase M (CPM)

Cathepsin S (CTSS)

Cell adhesion molecule 3 (CADM3)

Cell surface glycoprotein CD200 receptor 1 (CD200R1)

CMRF35-like molecule 1 (CLM-1)

CMRF35-like molecule 6 (CLM-6)

Contactin-5 (CNTN5)

Cytotoxic and regulatory T-cell molecule (CRTAM)

Dickkopf-related protein 4 (Dkk-4)

Dipeptidyl peptidase 1 (CTSC)

Disintegrin and metalloproteinase domain-containing protein 22 (ADAM 22)

Disintegrin and metalloproteinase domain-containing protein 23 (ADAM 23)

Draxin (DRAXIN)

Ephrin type-B receptor 6 (EPHB6)

Ephrin-A4 (EFNA4)

Epithelial discoidin domain-containing receptor 1 (DDR1)

Ezrin (EZR)

Fc receptor-like protein 2 (FcRL2)

Galectin-8 (gal-8)

GDNF family receptor alpha-1 (GFR-alpha-1)

GDNF family receptor alpha-3 (GDNFR-alpha-3)

Glial cell line-derived neurotrophic factor (GDNF)

Glypican-5 (GCP5)

Granulocyte Colony-Stimulating Factor (G-CSF)

Granulocyte-macrophage colony-stimulating factor receptor subunit alpha (GM-CSF-

R-alpha)

Granzyme A (GZMA)

Growth/differentiation factor 8 (GDF-8)

Hydroxyacylglutathione hydrolase, mitochondrial (HAGH)

Interleukin-12 subunit beta, Interleukin-12 subunit alpha (IL-12B, IL-12A)

Interleukin-5 receptor subunit alpha (IL-5R-alpha)

Junctional adhesion molecule B (JAM-B)

Kynureninase (KYNU)

Latexin (LXN)

Layilin (LAYN)

Leucine-rich repeat transmembrane protein FLRT2 (FLRT2)

Leukocyte-associated immunoglobulin-like receptor 2 (LAIR-2)

Linker for activation of T-cells family member 1 (LAT)

Lysosome membrane protein 2 (SCARB2)

Macrophage scavenger receptor types I and II (MSR1)

MAM domain-containing glycosylphosphatidylinositol anchor protein 1 (MDGA1)

Matrilin-3 (MATN3)

Mesencephalic astrocyte-derived neurotrophic factor (MANF)

N-acylethanolamine-hydrolyzing acid amidase (NAAA)

Neprilysin (NEP)

Netrin receptor UNC5C (UNC5C)

Neuroblastoma suppressor of tumorigenicity 1 (NBL1)

Neurocan core protein (NCAN)

Neuronal cell adhesion molecule (Nr-CAM)

Neuropilin-2 (NRP2)

Neutral ceramidase (N-CDase)

Nicotinamide/nicotinic acid mononucleotide adenylyltransferase 1 (NMNAT1)

NKG2D ligand 2 (N2DL-2)

NT-3 growth factor receptor (NTRK3)

OX-2 membrane glycoprotein (CD200)

Platelet-derived growth factor receptor alpha (PDGF-R-alpha)

Plexin-B1 (PLXNB1)

Plexin-B3 (PLXNB3)

Poliovirus receptor (PVR)

Protogenin (PRTG)

R-spondin-1 (RSPO1)

Repulsive guidance molecule A (RGMA)

RGM domain family member B (RGMB)

Roundabout homolog 2 (ROBO2)

Scavenger receptor class A member 5 (SCARA5)

Scavenger receptor class F member 2 (SCARF2)

Secreted frizzled-related protein 3 (sFRP-3)

Serine/threonine-protein kinase receptor R3 (SKR3)

Sialic acid-binding Ig-like lectin 9 (Siglec-9)

Sialoadhesin (SIGLEC1)

SPARC-related modular calcium-binding protein 2 (SMOC2)

Sphingomyelin phosphodiesterase (SMPD1)

Tenascin-R (TN-R)

Testican-1 (SPOCK1)

Thy-1 membrane glycoprotein (THY 1)

Transmembrane protease serine 5 (TMPRSS5)

Tumor necrosis factor receptor superfamily member 12A (TNFRSF12A)

Tumor necrosis factor receptor superfamily member 21 (TNFRSF21)

Tumor necrosis factor receptor superfamily member 27 (EDA2R)

WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 1

(WFIKKN1)

<u>Genetic Association study</u>

The adult cohort used in this part of the study had previously had genome-wide genotyping array performed using Illumina's Infinium MEGA chip (Multi Ethnic Genotyping Array). Genotyping data were quality controlled and imputed to 1000 Genomes phase 3 reference set using the online Michigan server. 359 individuals with SCD had genotyping as well as MRI data available, and were classified as having SCI or not; patients with large vessel vasculopathy, overt stroke or no brain MRI were excluded. We performed a mixed modeling analysis incorporating a genetic relatedness matrix to take account of relatedness (including population stratification), sex, age at MRI, and sickle genotype as fixed covariates. Each gene analysed contained multiple genetic variants; we corrected for multiple testing after quantifying the linkage disequilibrium within each region and used this to calculate appropriate significance levels for that region