Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome).

Pia Ostergaard[1,13], Michael A Simpson[2,13], Fiona C Connell[3], Colin G Steward[4], Glen Brice[5], Wesley J Woollard[2], Tatjana Kilo[6], Sarah Smithson[7], Peter Lunt[7], Victoria A Murday[8], Shirley Hodgson[5], Russell Keenan[9], Daniela T Pilz[10], Ines Martinez-Corral[11], Taija Makinen[11], Peter S Mortimer[12], Steve Jeffery[1], Richard C Trembath[2], Sahar Mansour[5]

1. Medical Genetics Unit, Biomedical sciences, St George's University of London, London, SW16 2BY, United Kingdom

2. Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Hospital, London, SE1 9RT, United Kingdom

3. Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, Guy's Hospital, London, SE1 9RT, United Kingdom

4. Bone Marrow Transplant Unit, Royal Hospital for Children, Bristol, BS2 8BJ, United Kingdom

5. South West Thames Regional Genetics Service, St. George's University of London, London, SW17 0RE, United Kingdom

6. Haematology Unit, The Children's Hospital at Westmead, Sydney, 2145, Australia

7. Department of Clinical Genetics, St Michael's Hospital, Bristol, BS2 8EG, United Kingdom

8. Department of Clinical Genetics, Yorkhill Hospital, Glasgow, G3 8SJ, United Kingdom

9. Department of Paediatric Haematology, Alderhey Children's Hospital, Liverpool, L12 2AP, United Kingdom

10. Institute of Medical Genetics, University Hospital of Wales, Cardiff , CF14 4XW, United Kingdom

11. Lymphatic Development Laboratory, Cancer Research UK London Research Institute, London, WC2A 3LY, United Kingdom

12. Department of Cardiac and Vascular Sciences, St George's University of London, London, SW17 0RE, United Kingdom

13. These authors contributed equally to this work

Correspondence to be addressed to smansour@sgul.ac.uk

Author Contributions

Jointly supervised research: PSM, SJ, RCT, SM

Conceived and designed the experiments: PO, MAS, FCC, GB, PSM, SJ, RCT, SM

Performed the experiments: PO, MAS, WJW, IC-M, TM

Performed statistical analysis: PO, MAS, TM

Analysed the data: PO, MAS, TM

Contributed reagents/materials/analysis tools: MAS, FCC, CS, GB, TK, SS, PL, VAM, SH,RK, DP, PSM, SJ, RCT, SM

Wrote the paper: MAS, SJ, RCT, SM

We report an allelic series of seven mutations in *GATA2* underlying Emberger syndrome, an autosomal dominant primary lymphedema associated with a predisposition to acute myeloid leukemia (AML). GATA2 is a transcription factor that plays an essential role in gene regulation during vascular development and hematopoietic differentiation. These findings indicate that haploinsufficiency of *GATA2* underlies primary lymphedema and predisposes to acute myeloid leukemia in this syndrome.

The co-occurrence of primary lymphedema with myelodysplasia progressing to acute myeloid leukemia (AML) termed Emberger syndrome (MIM:*pending*) has been reported as a sporadic disorder but also in a limited number of kindreds[1]. The syndrome segregates as an autosomal dominant trait with incomplete penetrance, and additional features may include abnormalities in the lymphocyte subsets, specifically low CD4/CD8 ratio, immune dysfunction as evidenced by the occurrence of severe and widespread cutaneous warts, and sensorineural deafness. Lymphedema consequent upon functional lymphatic hypoplasia, is confined to one or both of the lower limbs and genitalia, and predates the onset of hematological abnormalities. Karyotype anomalies associated with myelodysplasia include monosomy of chromosome 7, and transformation to AML is often rapid.

AML is a heterogeneous hematological malignancy and recent molecular genetic studies have identified recurrent somatic mutations in *DNMT3A, NPM1, FLT3, TET3, IDH1* and *IDH2* together with fusion proteins such as PML –RARX and CBFB –MYH11 each contributing to clonal development of subsets of the disease. Inherited predisposition to AML appears to be rare, having been described in a very few kindreds beyond disorders of primary bone marrow failure and inherited defects of DNA repair[2]. However, the delineation of the molecular genetic basis of familial forms of AML provides significant insight into the molecular mechanisms underlying myelodysplasia and leukemic transformation[2]. Germline *RUNX1* haploinsufficiency underlies an autosomal dominant familial platelet disorder with propensity to myeloid malignancy (MIM 601399)[3] and inherited mutations in *CEBPA* the gene encoding the transcription factor CCAAT/enhancer binding protein α (C/EBPα) has also been identified in a small number of families with AML[4]. Both RUNX1 and C/EBPα are transcription factors central to the development of normal hematopoiesis and somatic mutations in both genes have been identified in sporadic AML[5,6]. Taken together, these findings demonstrate the critical role of dysregulation of transcription control in hematological malignancies and AML in particular.

We sought to identify disease causing alleles in Emberger syndrome and recruited seven unrelated affected probands to this study. Two had a family history of primary lymphedema and/or AML and five were sporadic occurrences of the disease (**Table 1**). We recruited additional family members from the two affected kindreds (**Supplementary** Figure 1). Clinical investigation of familial cases provided significant insight into the phenotypic variability of Emberger syndrome; within one family (Emb-01). This ranged from hydrops fetalis evolving to severe bilateral lower limb lymphedema (**Figure 1A and B**) with genital involvement and development of AML at 11 years of age (Emb-01 III-1) to an obligate carrier (her father, Emb-01 II-1) with minimal edema and a normal lymphocyte count but low CD4/CD8 ratio (case reports of Emb-01 presented as case 1 in reference 1). We undertook whole exome sequencing of three unrelated affected individuals; one sporadic case and two subjects from kindreds with the disorder (**Table 1**). Whole exome capture was performed by in-solution hybridisation followed by massively parallel sequencing (**Supplementary** Methods). Over 6.8 gigabases of sequence was generated for each subject, such that >80% of the coding bases of the Gencode defined exome were represented by at least 20 reads (**Supplementary** Table 1). Single nucleotide substitutions and small insertion deletion variants were identified using our in house variant calling pipeline[7] (**Supplementary** Methods and **Supplementary** Table 2). Analysis of the exome variant profiles was performed under a model of a rare autosomal dominant disorder, requiring at least one previously unobserved heterozygous nonsynonymous or splice site substitution or a coding insertion or deletion in the same gene in all three individuals, a process that highlighted *GATA2* as the only candidate gene matching these criteria (**Supplementary** Methods and **Supplementary** Table 3). The three *GATA2* variants are each predicted to lead to premature termination of the protein product with a high likelihood of significant functional impact (**Figure 1C**). Each of the three mutations was confirmed by Sanger sequencing and assessed in all available relatives. The two variants identified in the familial cases were demonstrated to be transmitted across generations between affected subjects within the kindred (**Table 1**).

We next addressed the hypothesis that further mutant *GATA2* alleles may provide the molecular genetic basis of additional cases of Emberger syndrome. We performed Sanger sequencing of the five *GATA2* coding exons and their associated splice sites in four additional independent cases (**Supplementary** Methods). We identified novel heterozygous genetic variants in all four subjects (**Figure 1C** and **Table 1**). Two are predicted to lead to premature termination and hence are likely to have significant functional impact. The remaining two lead to substitution of conserved residues in a known and critical functional domain (**Supplementary** Methods and **Supplementary** Figure 2). The GATA family of transcription factors comprises six proteins in humans. All have two adjacent C_4 zinc

finger DNA-binding domains each containing four cysteine residues with the C373R and R361L amino acid substitutions both residing within the second C_4 zinc-finger domain of the protein. This family of molecules controls the development of a wide range of tissues by activating and repressing transcription through the zinc finger mediated binding to the GATA consensus sequence motif (A/T)GATA(A/G). In total we identified seven independent novel *GATA2* variants in 13 individuals (**Table 1**). None of the identified variants were detected in 300 unrelated control chromosomes. Each variant is predicted to have substantial impact upon the function of GATA2 suggesting that dysregulation of gene expression controlled by this transcription factor is responsible for Emberger syndrome and inherited predisposition to AML. The spectrum of clinical findings across the 13 mutation carriers confirms the variable presentation of the disorder (**Table 1**).

GATA2 is expressed in hematopoietic stem cells, multipotent hematopoietic progenitors and beyond the hematopoietic system. Homozygous *Gata2* knock-out mice die during mid gestation due to severe anaemia with levels of myeloid-erythroid progenitor cells significantly reduced in comparison to wild-type controls[8]. Analysis of heterozygous mice haploinsufficient for *Gata2* revealed disruption of hematopoietic stem cell homeostasis within the granular macrophage progenitor compartment, which has previously been shown to be vulnerable to leukemic transformation[9].

GATA transcription factors have previously been implicated in human cancers (reviewed in reference 10). Indeed, somatic mutation of *GATA2* has been identified in samples from patients with chronic myeloid leukemia undergoing blast transformation[11]. The mutations we present here provide the first description of germline defects in *GATA2* underlying an inherited predisposition to AML, findings that are supported by the recent identification of a microdeletion at 3q21, encompassing 36 genes including *GATA2,* in a patient with multiple congenital abnormalities and development of myelodysplasia with monosomy 7 at the age of 11 years[12]. Also of interest and relevant to our findings, GATA2 has been shown to form protein complexes C/EBPα[13], implicating mutation of either of the genes that encode for these transcription factors in inherited predisposition to AML.

A limited number of genes have been shown to be critical for the development and maintenance of a functional lymphatic system in humans with disruption leading to primary lymphedema. These include loss of function mutations of the tyrosine kinase domain of *VEGFR3* and defects in *FOXC2, SOX18, CCBE1* and *GJC2. GATA2* is known to be expressed in lymphatic, vascular and endocardial endothelial cells[14]. However, neither heterozygous nor homozygous knock-out mice were reported with any overt defect of the vasculature, which may suggest a level of functional redundancy of Gata2 to other Gata family members in these tissues in the mouse. We also note the described

function of Gata2 in vestibular morphogenesis and growth of the semicircular canals[15], the relationship of these findings with the development of sensorineural deafness in three of the 13 mutation carriers requires further investigation.

Our findings demonstrate the critical role of *GATA2* in the development and maintenance of the lymphatics and hematopoietic system. Further investigation is required to delineate fully the molecular and cellular mechanisms that contribute to the variable phenotypic expression of Emberger syndrome including predisposition to AML. The identification of *GATA2* as the gene responsible for this syndrome will facilitate the diagnosis and monitoring of patients with primary lymphoedema who are at increased risk of developing this life-threatening hematological malignancy.

References

1. Mansour S. et al. *Am J Med Genet* **152A**, 2287-2296 (2010).
2. Owen C., Barnett M. and Fitzgibbon J. *Brit J Haematol* **140**, 123-132 (2008).
3. Song W-J. et al. *Nat Genet* **23**, 166-175 (1999).
4. Smith M. L. et al. *N Eng J Med* **351**, 2403-2407 (2004).
5. Preudhomme C. *et al. Blood* **96**, 2862-2869 (2000).
6. Pabst T. *et al. Nat Genet* **27**, 263-270 (2001).
7. Ostergaard P *et al. J Med Genet* **48**, 251-255 (2011).
8. Tsai F-Y. *et al. Nature* **371,** 221-226 (1994).
9. Rodrigues N.P. *et al Blood* **112,** 4862-4873 (2008).
10. Zheng R. and Blobel G.A. *Genes and Cancer* epub before print doi:10.1177/1947601911404223
11. Zhang S-J *et al. PNAS* **105,** 2076-2081 (2008).
12. Callier P. *et al. Am J Med Genet* **149A**, 1323–1326 (2009).
13. Tong Q. *et al. Mol Cell Biol* **25**, 706-715 (2005)
14. Khandekar M. *et al. Development* **134,** 1703-1712 (2007).
15. Haugas M. *et al Dev Dyn* **239** 2452–2469 (2010)

URLs

The GENECODE Project

http://www.gencodegenes.org/

Accession Numbers

NM_032638

Figure 1: A. Bilateral lower limb lymphedema in patient Emb-01 III-1. B. Four limb lymphoscintigraphy performed in subject Emb-01 III-1 showing no significant main tract filling of the lower limbs at 2 hours. C. Location of the seven identified mutations with respect to the genomic organisation of the *GATA2* gene (upper panel) and GATA2 protein domain structure (lower panel).

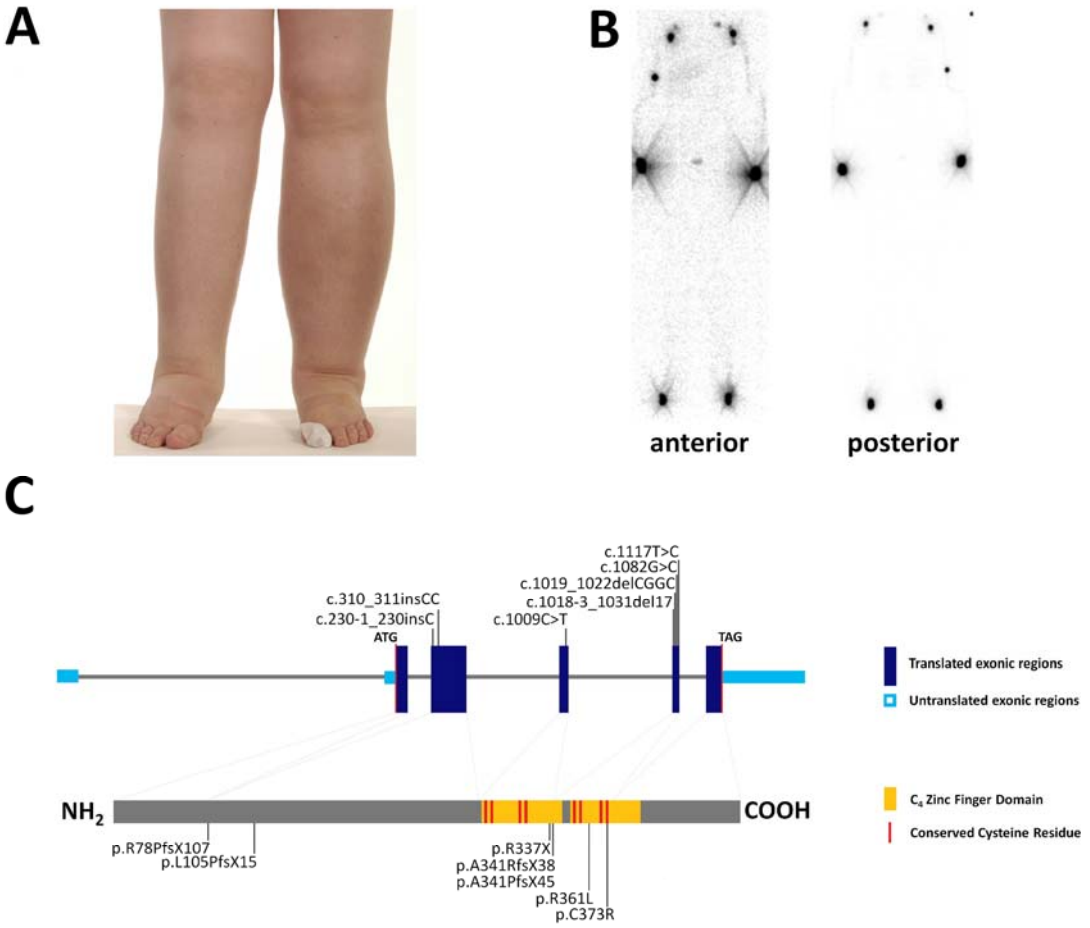

Table1: Clinical and genetic findings in Emberger syndrome

Pedigree	Individual	Gender	Familial /Sporadic	Lymphedema (age at diagnosis)	Hematological abnormalities (age at diagnosis)	Additional clinical features	Nucleotide variant	Predicted protein alteration
Emb-01	**II-1***	Male	Familial	Minimal edema both feet	Low CD4/CD8 ratio	None	c.310_311insCC	p.L105PfsX15
	II-4	Female	Familial	Bilateral lower limb (44 years[#])	Myelodysplasia (50 years) AML (53 years)	Died (53 years) Cutaneous warts with malignant transformation to anogenital dysplasia	c.310_311insCC	p.L105PfsX15
	III-1	Female	Familial	Hydrops fetalis, Bilateral lower limb and genital (birth)	Myelodysplasia (11 years) AML (11 years)	None	c.310_311insCC	p.L105PfsX15
	III-3	Female	Familial	None	Myelodysplasia (9 years) AML (9 years)	Died (9 years)	c.310_311insCC	p.L105PfsX15
	III-4	Male	Familial	Left lower limb and genital(14 years)	Low CD4/CD8 ratio	Unilateral ptosis	c.310_311insCC	p.L105PfsX15
Emb-02	**I-2**	Female	Familial	Minimal edema both feet	None	None	c.230-1_230insC	p.R78PfsX107
	II-1	Female	Familial	Left lower limb (16 years)	Immature bone marrow[+]	Bilateral cleft lip and palate	c.230-1_230insC	p.R78PfsX107
	II-2*	Male	Familial	Left lower limb and genital (16 years)	Myelodysplasia (17 years) AML (17 years)	Died (17 years)	c.230-1_230insC	p.R78PfsX107
Emb-03	**I-1***	Female	Sporadic	Bilateral lower limb and genital(6 years)	Myelodysplasia (12 years) AML (12 years)	None	c.1009C>T	p.R337X
Emb-04	**I-1**	Male	Sporadic	Left lower limb and genital (6 years)	Myelodysplasia (11 years)	Sensorineural hearing loss	c.1019_1022delCGGC	p.A341PfsX45
Emb-05	**I-1**	Female	Sporadic	Bilateral lower limb (birth)	Low CD4/CD8 ratio	Cutaneous warts with malignant transformation to anogenital dysplasia Sensorineural hearing loss	c.1018-3_1031del17	p.A341RfsX38
Emb-06	**I-1**	Male	Sporadic	Bilateral lower limb and genital (8 years)	Myelodysplasia (16 years)	Died (16 years) Persistent warts on fingers	c.1117T>C	p.C373R
Emb-07	**I-1**	Male	Sporadic	Bilateral lower limb and genital (10 years)	Low CD4/CD8 ratio	Cutaneous warts with malignant transformation to anogenital dysplasia Sensorineural hearing loss	c.1082G>C	p.R361L

*individuals exome sequenced in the primary analysis, #onset of lymphedema post surgery, +Monosomy 7 detected in bone marrow aspirate

Acknowledgements

This study has been made possible by the commitment and courage of the families concerned, we are indebted to them. This work was supported by the British Heart Foundation (BHF) grants RG/08/006/25302, PG/10/58/28477 and FS/06/063/21445. The authors also acknowledge support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. This research utilised the resources of the Biomics Unit at St George's University of London.

Competing Financial Interests

The authors declare no competing financial interests.