

# Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency

Rachel Emma Dickinson,<sup>1</sup> Helen Griffin,<sup>2</sup> Venetia Bigley,<sup>1,3</sup> Louise N. Reynard,<sup>1</sup> Rafiqul Hussain,<sup>2</sup> Muzlifah Haniffa,<sup>1,3</sup> Jeremy H. Lakey,<sup>4</sup> Thahira Rahman,<sup>2</sup> Xiao-Nong Wang,<sup>1</sup> Naomi McGovern,<sup>1</sup> Sarah Pagan,<sup>1</sup> Sharon Cookson,<sup>1</sup> David McDonald,<sup>1</sup> Ignatius Chua,<sup>5</sup> Jonathan Wallis,<sup>3</sup> Andrew Cant,<sup>1,3</sup> Michael Wright,<sup>2,3</sup> Bernard Keavney,<sup>2</sup> Patrick F. Chinnery,<sup>2</sup> John Loughlin,<sup>1</sup> Sophie Hambleton,<sup>1,3</sup> Mauro Santibanez-Koref,<sup>2</sup> and Matthew Collin<sup>1,3</sup>

<sup>1</sup>Institute of Cellular Medicine and <sup>2</sup>Institute of Genetic Medicine, International Centre for Life, Newcastle University, Newcastle upon Tyne, United Kingdom; <sup>3</sup>Newcastle upon Tyne Hospitals National Health Service Foundation Trust, Newcastle upon Tyne, United Kingdom; <sup>4</sup>Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, United Kingdom; and <sup>5</sup>University College London Centre for Immunodeficiency, Royal Free Hospital, London, United Kingdom

The human syndrome of dendritic cell, monocyte, B and natural killer lymphoid deficiency presents as a sporadic or autosomal dominant trait causing susceptibility to mycobacterial and other infections, predisposition to myelodysplasia and leukemia, and, in some cases, pulmonary alveolar proteinosis. Seeking a genetic cause, we sequenced the exomes of

4 unrelated persons, 3 with sporadic disease, looking for novel, heterozygous, and probably deleterious variants. A number of genes harbored novel variants in person, but only one gene, GATA2, was mutated in all 4 persons. Each person harbored a different mutation, but all were predicted to be highly deleterious and to cause loss or mutation of the C-terminal

zinc finger domain. Because GATA2 is the only common mutated gene in 4 unrelated persons, it is highly probable to be the cause of dendritic cell, monocyte, B, and natural killer lymphoid deficiency. This disorder therefore constitutes a new genetic form of heritable immunodeficiency and leukemic transformation. (*Blood*. 2011;118(10):2656-2658)

## Introduction

We recently described an immunodeficiency syndrome that involved loss of dendritic cells (DCs), monocytes, and B and natural killer cells (DCML deficiency).<sup>1</sup> A similar disorder of monocytopenia with *Mycobacterium avium* complex infection has been described as “monoMAC.”<sup>2,3</sup> Patients typically present in young adulthood with disseminated mycobacterial and papilloma virus infection occasionally complicated by pulmonary alveolar proteinosis. Left untreated, their lives are shortened by refractory infection, respiratory failure, and leukemic transformation.

Autosomal dominant inheritance among approximately one-half of cases strongly suggests a monoallelic genetic disorder.<sup>2</sup> We therefore sequenced exomes to identify novel heterozygous mutations in these patients. Assuming a single gene cause, we reasoned that sequencing 4 unrelated persons would exclude nearly all private but non-disease-associated mutations. We also compared the sequence of constitutional and hematopoietic DNA from 2 persons, to exclude somatic loss of heterozygosity in the hematopoietic stem cell compartment as a potential biallelic mechanism of autosomal dominant inheritance.

## Methods

Patients with DCML deficiency were identified as previously described.<sup>1</sup> All clinical samples and data were obtained with local ethical permission from the Newcastle and North Tyneside Research Ethics Committee, which approved the study. Genomic DNA was extracted with the QIAamp DNA Mini-Kit (QIAGEN); fragmented to 150-200 bp with the use of the

Adaptive Focused Acoustics (Covaris); then end-repaired, adenylated, and ligated to adapters (Illumina Paired-End Sample Preparation kit). Ligated libraries were hybridized with whole-exome baits that covered 27 184 genes (Agilent SureSelect Human All Exon Kit Version 2) with modifications for the SureSelect Target Enrichment System for Illumina Paired-end Sequencing Library (Version 2.0.1). The captured fragments were purified and clonally amplified then sequenced on 2 lanes of an Illumina Genome Analyser IIx with the use of 75 bp paired-end reads.

Sequence was aligned to the human reference genome (UCSC hg19) with the use of Bowtie<sup>4</sup> for single-base variants and BWA<sup>5</sup> for indels, then reformatted with the use of SAMtools.<sup>6</sup> Eighty-seven percent of exon target sequence was covered by > 10 reads with an average coverage of 90 reads. Single base variants were identified with Varscan,<sup>7</sup> and indels were identified with Dindel.<sup>8</sup> Lists of on-target variants were filtered against dbSNP131 (1094 persons) and the exome sequences of 10 unrelated and unaffected persons. Putative disease-causing mutations were identified with MutationTaster.<sup>9</sup> GATA2 sequence was covered by 12-96 reads for novel variants and 16-86 reads for reference sequence. Novel GATA2 variants were confirmed by PCR amplification (supplemental Table 1, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article) and Sanger sequencing and were not found in 11 unaffected first-degree relatives of the patients. An x-ray-derived structure of GATA-binding factor (GATA) zinc finger was used to model the mutations.<sup>10</sup> Mutants R398W and T354M were created in Swiss PDB Modeller,<sup>11</sup> and the resulting structures were displayed with Pymol (Delano Scientific).

## Results and discussion

Exome sequencing identified between 853 and 1968 novel variants in each person (Table 1). Of these, 190-370 variants were predicted

Submitted June 10, 2011; accepted June 21, 2011. Prepublished online as *Blood* First Edition paper, July 15, 2011; DOI 10.1182/blood-2011-06-360313.

An Inside *Blood* analysis of this article appears at the front of this issue.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

© 2011 by The American Society of Hematology

**Table 1. Frequency and distribution of novel variants between persons with DCML deficiency**

Patient* (tissue)	No. of novel variants single base + indels	No. of predicted disease-causing variants single base + indels	No. of variant genes shared by 2 of 4 subjects single base + indels	Identity of variant genes shared by 3 of 4 subjects (indel indicated)	Identity of variant gene shared by 4 of 4 subjects
1 (fibroblasts)	803 + 160	172 + 25	13 + 3	ADAMTS3; DNMBP; <i>GATA2</i> ; (indel) PLEC; RNF160	<i>GATA2</i>
2 (fibroblasts)	733 + 120	172 + 18	18 + 2	ADAMTS3; DNMBP; <i>GATA2</i> ; PLEC; RNF160	<i>GATA2</i>
3 (fibroblasts)	1046 + 272	243 + 40	15 + 2	ADAMTS3; <i>GATA2</i> ; <i>GATA2</i>	
(neutrophils)	774 + 113	213 + 29			
4 (fibroblasts)	1827 + 115	342 + 22	18 + 3	DNMBP; <i>GATA2</i> ; PLEC; RNF160	<i>GATA-2</i>
(neutrophils)	1871 + 97	348 + 22			

\*Numbering as in Bigley et al.<sup>1</sup>

to be disease causing by MutationTaster.<sup>9</sup> Any 2 persons shared 17-21 genes harboring novel variants and any 3 persons shared 2-6 genes. All shared variant genes contained only heterozygous changes. Only one gene with novel disease-causing variants was found in all 4 patients: *GATA2*. Identification of a single variant gene in 4 unrelated persons with a common disease phenotype strongly implicates this gene in the pathogenesis of the disorder. Given that 190-370 deleterious variants were detected in 27 184 genes, the chance of observing a novel deleterious variant of any gene is 0.6%-1.3%. The probability of 4 unrelated persons harboring a single variant gene in common is  $< 1 \times 10^{-3}$ . Each patient had a unique mutation. Neutrophil DNA sequenced in 2 patients also had heterozygous mutation of *GATA-2*, excluding somatically acquired loss of heterozygosity as a mechanism of pathogenesis.<sup>12</sup> These results illustrate the power of exome sequencing to identify disease-causing mutations in a small number of unrelated and sporadic cases of a genetic disorder.

*GATA-2* is a transcription factor required for stem cell homeostasis.<sup>13</sup> The protein contains 2 highly conserved zinc finger domains that mediate protein-DNA and protein-protein interactions. Examination of the mutations in *GATA2* showed a base insertion predicted to cause a frameshift after G200 and subsequent premature stop codon, 2 single nucleotide substitutions causing amino acid changes T354M and R398W, and a splice acceptor site mutation predicted to cause skipping of exon 5 and a 42-aa deletion (Table 2). The frameshift and splice acceptor mutations are expected to cause complete loss of the C terminal zinc finger and are probably null mutations, anticipated to cause DCML deficiency through haploinsufficiency of *GATA-2*. The amino acid substitutions probably also introduce significant structural variations in the C terminal zinc finger (Figure 1). Substitution of methionine at 354 is predicted to disrupt the conserved threonine string of the zinc finger domain, and replacement of positively charged arginine by tryptophan at the C terminus of the zinc finger probably prevents a critical interaction with the minor groove of DNA. One of these mutations, T354M, has recently been reported as a cause of autosomal-dominant familial myelodysplastic syndrome/acute myeloid leukemia. In this case, in vitro studies indicate dominant-

negative function of the mutant protein in certain situations.<sup>14</sup> In summary, both haploinsufficiency and dominant-negative loss of *GATA-2* function are potential mechanisms of pathogenesis in DCML deficiency.

Investigation of the stem cell compartment of patients with DCML deficiency highlighted the absence of multi-lymphoid progenitors and decreased granulocyte-macrophage progenitors in the face of highly elevated flt-3 ligand.<sup>1,15</sup> This situation is in marked contrast to all other DC deficiency states, including IRF8 mutation in humans, which are associated with a myeloproliferative response.<sup>16</sup> The lack of myeloproliferation in DCML deficiency is in keeping with an intrinsic stem cell defect, as might be predicted to occur with *GATA2* mutation. Further work will be required to elucidate the preferential drop-out of mononuclear cell precursors that gives rise to immunodeficiency early in the disease.<sup>1,2</sup> In mice with haploinsufficiency of *Gata-2* because of heterozygous truncation of the C-terminal zinc finger domain, the region primarily affected in the persons described, specific defects of the granulocyte/macrophage progenitor pool have been reported.<sup>17</sup> Intriguingly, there is also selective loss of monocytes and lymphoid cells but preservation of granulopoiesis when *Gata2* heterozygous stem cells are compared with wild-type stem cells in competitive transplantation assays.<sup>13</sup> These results indicate that haploinsufficiency of *GATA-2* modifies hematopoietic differentiation and could be sufficient to cause autosomal-dominant DCML deficiency in humans. Numerous downstream targets of *GATA-2* are involved in terminal hematopoietic differentiation. Of note, *GATA-2* interacts with the myeloid transcription factor PU.1,<sup>18</sup> which is required to drive monocyte colony-stimulating factor receptor and Flt-3 expression, both essential growth factor receptors in the development of DCs and monocytes.<sup>19</sup> We previously found a marked reduction in Flt-3 expression by CD34<sup>+</sup> cells in DCML deficiency.<sup>1</sup> The regulation of phagocytosis by *GATA-2* may also explain loss of function in alveolar macrophages, leading to pulmonary alveolar proteinosis in some patients.<sup>20</sup>

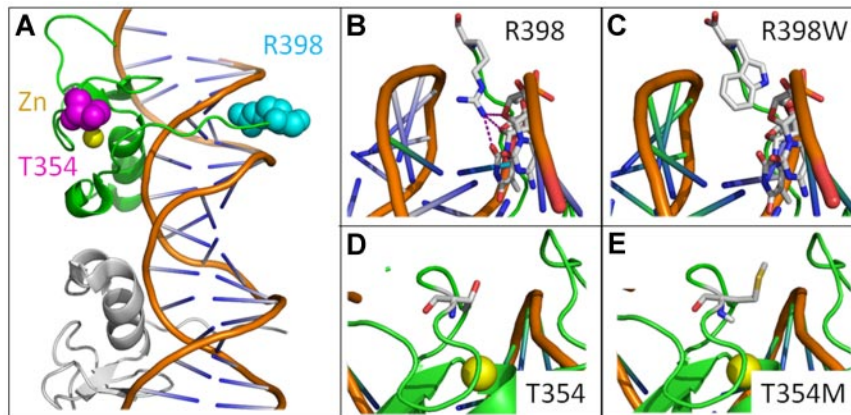
Mutations of *GATA-2* also cause familial myelodysplastic syndrome/acute myeloid leukemia.<sup>14</sup> Although DCML deficiency can exist for many years without tri-lineage dysplasia, the evolution of progressive pancytopenia, and acquired cytogenetic abnormalities, including monosomy 7 and trisomy 8, have been highlighted in a recent study.<sup>3</sup> This provides a unified view of *GATA-2* mutation as a cause of both immunodeficiency and myeloid malignancy through its unique function in maintaining both mononuclear cell differentiation and stem cell homeostasis.

Note added in proof: Coincident with submission of our paper, Holland's group published a paper reporting the association of *GATA2* mutations with the MonoMac syndrome.<sup>21</sup>

**Table 2. Mutations of *GATA2* and predicted amino acid changes**

Subject	DNA change*	Codon
1	c.599_600insG	G200fs
2	c.1061 C > T	T354M
3	c.1192 C > T	R398W
4	c.1018-1 G > T	Δ340-381

\*Numbering relative to adenine in the ATG start codon of *GATA2* (NM\_001145661.1) and the first methionine of the *GATA2* protein (NP\_116027.2).



**Figure 1. Mutations in GATA zinc finger domain.** GATA zinc finger domains bind DNA in 2 conformations the adjacent and opposite forms. Here, we use the adjacent structure (3DFV).<sup>10</sup> (A) Illustrated view of single GATA zinc finger (green) bound to DNA with adjacent GATA molecule in gray. R398 (cyan), T354 (magenta), and zinc ion (yellow) shown in space filling representation. (B) Wild-type R398 shown inserting into minor groove and possible H bonds with DNA bases. (C) R398W mutation modeled. (D) Wild-type T354. (E) T354M mutation modeled showing increased side chain size.

## Acknowledgments

The authors thank Steven Holland for giving access to unpublished data relating to his laboratory's study on GATA-2 mutation,<sup>21</sup> which was received while this manuscript was in preparation.

This work was supported by Leukaemia and Lymphoma Research, UK (R.E.D. and M.C.); Medical Research Council, United Kingdom (V.B., P.F.C., and S.H.); the Wellcome Trust (M.H. and P.F.C.); British Heart Foundation (B.K.); and by the National Institute of Health Research Biomedical Research Center at Newcastle Hospitals NHS Foundation Trust, The Newcastle Health Care Charity, and the Newcastle on Tyne Hospitals NHS Charity.

## Authorship

Contribution: R.E.D. performed sequencing, analyzed data, and wrote the manuscript; H.G. designed and performed bioinformatic

analysis and wrote the manuscript; V.B. identified the phenotype of subjects, designed the study, and wrote the manuscript; L.N.R. processed samples and commented on the manuscript; R.H. performed sequencing and commented on the manuscript; M.H. identified the phenotype of patients and commented on the manuscript; J.H.L. performed structural modeling; T.R. performed sequencing and wrote the manuscript; X.-N.W., N.Mc.G., and D.Mc.D. processed samples; S.P. and S.C. provided technical assistance; I.C., J.W., A.C., and S.H. identified patients; M.W., B.K., P.F.C., J.L., and S.H. analyzed data and commented on manuscript; M.S.-K. performed bioinformatic analysis and commented on manuscript; and M.C. designed the study, analyzed data, and wrote the manuscript

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Matthew Collin, Institute of Cellular Medicine, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom; e-mail: matthew.collin@ncl.ac.uk.

## References

- Bigley V, Haniffa M, Doulatov S, et al. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J Exp Med*. 2011;208(2):227-234.
- Vinh DC, Patel SY, Uzel G, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood*. 2010;115(8):1519-1529.
- Calvo KR, Vinh DC, Maric I, et al. Myelodysplasia in autosomal dominant and sporadic monocytopenia immunodeficiency syndrome: diagnostic features and clinical implications [published online ahead of print April 20, 2011]. *Haematologica*. doi:10.3324/haematol.2011.041152.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
- Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
- Koboldt DC, Chen K, Wylie T, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283-2285.
- Albers CA, Lunter G, Macarthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. *Genome Res*. 2011;21(6):961-973.
- Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575-576.
- Bates DL, Chen Y, Kim G, Guo L, Chen L. Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J Mol Biol*. 2008;381(5):1292-1306.
- Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18(15):2714-2723.
- O'Keefe C, McDevitt MA, Maciejewski JP. Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood*. 2010;115(14):2731-2739.
- Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*. 2005;106(2):477-484.
- Scott HS, Hahn CH, Carmichael CL, et al. GATA2 is a new predisposition gene for familial myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML) [abstract]. *Blood*. 2010;116(21):Abstract 3.
- Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol*. 2010;11(7):585-593.
- Hambleton S, Salem S, Bustamante J, et al. IRF8 Mutations and human dendritic-cell immunodeficiency. *N Engl J Med*. 2011;365(2):127-138.
- Rodrigues NP, Boyd AS, Fugazza C, et al. GATA-2 regulates granulocyte-macrophage progenitor cell function. *Blood*. 2008;112(13):4862-4873.
- Kitajima K, Tanaka M, Zheng J, et al. Redirecting differentiation of hematopoietic progenitors by a transcription factor, GATA-2. *Blood*. 2006;107(5):1857-1863.
- Carotta S, Dakic A, D'Amico A, et al. The transcription factor PU. 1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. *Immunity*. 2010;32(5):628-641.
- Lasbury ME, Tang X, Durant PJ, Lee CH. Effect of transcription factor GATA-2 on phagocytic activity of alveolar macrophages from *Pneumocystis carinii*-infected hosts. *Infect Immun*. 2003;71(9):4943-4952.
- Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011;118(10):2653-2655.