Rapid identification of mutations in GJC2 in primary lymphoedema using whole exome sequencing combined with linkage analysis with delineation of the phenotype

Pia Ostergaard,1 Michael A Simpson,2 Glen Brice,3 Sahar Mansour,3 Fiona C Connell,4 Alexandros Onoufrriadis,2 Anne H Child,5 Jae Hwang,1 Kamini Kalidas,1 Peter S Mortimer,6 Richard Trembath,2 Steve Jeffery1

ABSTRACT

Background Primary lymphoedema describes a chronic, frequently progressive, failure of lymphatic drainage. This disorder is frequently genetic in origin, and a multigenerational family in which eight individuals developed postnatal lymphoedema of all four limbs was ascertained from the joint Lymphoedema/Genetic clinic at St George’s Hospital.

Methods Linkage analysis was used to determine a locus, and exome sequencing was employed to look for causative variants.

Results Linkage analysis revealed cosegregation of a 16.1 Mb haplotype on chromosome 1q42 that contained 173 known or predicted genes. Whole exome sequencing in a single affected individual was undertaken, and the search for the causative variant focused to within the linkage interval. This approach revealed two novel non-synonymous single nucleotide substitutions within the chromosome 1 locus, in NVL and GJC2. NVL and GJC2 were sequenced in an additional cohort of individuals with a similar phenotype and non-synonymous variants were found in GJC2 in four additional families.

Conclusion This report demonstrates the power of exome sequencing efficiently applied to a traditional positional cloning pipeline in disease gene discovery, and suggests that the phenotype produced by GJC2 mutations is predominantly one of 4 limb lymphoedema.

INTRODUCTION

The accumulation of protein-rich fluid in the interstitial spaces results from an anatomical or functional defect in the lymphatic vessels and can occur as a non-syndromic Mendelian condition or as part of a more complex, syndromic disorder.1 Primary lymphoedema most commonly affects the lower limbs, but other body parts (upper limbs, face, and genitalia) can also be affected. Lymphoscintigraphy occasionally reveals lymphatic abnormalities in the upper limbs in individuals presenting with oedema of the lower limbs.3 However, individuals with clinical signs of primary lymphoedema in all four limbs are not commonly seen in the lymphoedema clinic. Connell et al2 reported 23 cases out of a cohort of 353 probands with late onset multisegmental primary lymphoedema affecting other body parts in addition to the lower limbs. A proportion of these had an autosomal dominant family history (F Connell, personal communication, 2010).

As with many other Mendelian traits, positional cloning by linkage mapping has proved to be a powerful approach in the identification of genes underlying primary lymphoedema. To date five genes have been shown to be mutated in disorders where lymphoedema is the major feature. Three genes were identified by linkage based positional cloning; FLT4 (VEGFR3) in Milroy disease,2–5 FOXC2 in lymphoedema distichiasis syndrome6 7 and SOX18 in the rare hypotrichosis–lymphoedema–telangiectasia syndrome.8 In addition, the human mutation of CCBE1 was identified in generalised lymphatic dysplasia/Hennekam syndrome9 10 because loss of function defects in ccbe1 in the zebra fish mutant ‘full of fluid’ (fof) highlighted the role of the CCBE1 protein product in lymphangiogenesis.11 Most recently, Ferrell et al1 identified mutations in GJC2 after differential transcript expression analysis highlighted a potential role of this gene in lymphoedema.12

The experimental route from linkage interval to causative mutation has traditionally been by Sanger sequencing of the coding regions and associated splice sites of the genes located within the delimiting recombination boundaries. While this approach has proved successful it is often limited by the size of the region of interest and the number of genes within it. Sequencing of more than a handful of genes can become a costly and time consuming process. However, recent developments in high throughput second generation sequencing and hybrid capture techniques15 now offer a more cost effective approach to this sequencing phase. Several recent investigations have shown the utility of customisable capture arrays for sequencing genes located within linkage intervals.16–17 However, improvements in capture efficiency with ‘in solution’ methods combined with increases in sequence yield now make whole exome sequencing an affordable alternative.

We present a large, multigenerational pedigree (figure 1) in which four limb primary lymphoedema is segregating in an autosomal dominant manner. Linkage analysis combined with whole exome sequencing rapidly identified two candidate genes with potential pathogenic variants co-segregating...
with the disease status. Subsequent analysis of these two genes in a cohort of additional individuals with primary lymphoedema confirmed that the identified defect in GJC2 underlies four limb lymphoedema in family I, and also provides evidence that mutations in GJC2 contribute to a significant proportion of autosomal dominant lymphoedema cases.

RESULTS

Clinical assessment of four limb lymphoedema in family I

Family I

Seven affected members of this family along with six at-risk relatives were clinically examined by one of the authors (GB). The inheritance pattern was consistent with autosomal dominance and there was no evidence of non-penetration. Expression was variable, ranging from mild, below knee swelling to severe, four limb swelling. Age of initial presentation of oedema ranged from 8–14 years. The two individuals in the family with the most severe swelling had suffered multiple episodes of cellulitis.

One individual had undergone isotope lymphoscintigraphy which was reported to show poor uptake in all four limbs, suggestive of distal hypoplasia. No members of this family had vein scan results available; however, one individual did report removal of the long saphenous vein in his 20s.

Linkage analysis

Genotyping of 6000 SNPs distributed throughout the genome (Illumina Human Linkage 12) in eight affected individuals and four unaffected relatives in family I (figure 1) revealed a 16.1 Mb chromosomal region, delimited by markers rs10494988 and rs1043909, co-segregating with the disease status. Linkage analysis of these data performed with Merlin, under a model of autosomal dominant inheritance with 99% penetrance and a 1% phenocopy rate, generated a maximum LOD of 2.94 within the chromosome 1 region.

Sequencing

The putative 16.1 Mb region on chromosome 1 contains 173 known or predicted genes. A strategy of whole exome sequencing by hybrid capture and second generation sequencing was employed. Sequencing of the CCDS defined exome was undertaken in a single affected individual from family I (individual III.10, figure 1). The efficiency of the hybrid capture was 73.4% with 51494434 of the 70164176 uniquely mapped sequence reads originating from the targeted exonic regions (75.1% to the targeted regions ±150 bp). This level of capture efficiency and sequence generation resulted in a mean coverage of the CCDS defined exome of 97×, with 91.7% of CCDS bases covered by at least 10 reads. Within the chromosome 1 linkage interval (rs10494988 to rs1043909), 20 candidate heterozygous non-synonymous variants were identified, of which 18 were excluded based on their presence in genomic variant databases. This left just two candidate heterozygous variants, in NVL (c.2236G>A; p.Val746Met), and GJC2 (c.143C>T; p.Ser48Leu). Subsequent genotyping of both variants in the entire family revealed that both variants co-segregated with the disease status. Neither change was seen in 150 control individuals.

Mutation validation

Initial investigation to establish which of the variants was pathogenic was undertaken by Sanger sequencing of coding exons and associated splice sites of NVL and GJC2 in a second family (family II, figure 2) also consistent with linkage to the same chromosome 1 region. Interestingly, the two variants

METHODS

Patients and controls

Patients were ascertained through the joint dermatology/genetics lymphoedema clinic at St George’s Hospital, as part of the ongoing UKCRN registered study into the genetic causes of primary lymphoedema. There is current ethical permission for this project. Patients were clinically assessed and where possible lymphoscintigram and venous duplex scans were performed. Use of a diagnostic algorithm18 allowed grouping of patients by phenotype. Initially, two probands from families I and II with lymphoedema of all four limbs and extensive positive family histories were identified for molecular analysis. Subsequently, other patients, from whom DNA samples were available, with similar and different primary lymphoedema phenotypes were studied (see Results for further details). All variants identified were assessed in a cohort of 150 control individuals of European ancestry.

SNP chip array and linkage analysis

DNA was extracted from peripheral blood using a standard chloroform ethanol procedure. Single nucleotide polymorphism (SNP) genotyping was undertaken on 12 individuals from family I. Processing of SNP microarrays (Illumina, Human Linkage-12 panel with 6090 SNP markers) was performed according to the manufacturer’s protocol. All genotype calls were checked for Mendelian inconsistencies and parametric LOD (logarithm of odds) scores calculated with MERLIN19 using an autosomal dominant model, 99% penetrance and 1% phenocopy rate.

Exome sequencing

Genomic DNA from the proband of family I was fragmented and enriched for exonic sequences with the Agilent SureSelect Whole Exome hybrid capture. The resulting enriched sequence library was sequenced with 76 bp paired end reads across two lanes of Illumina GAIIx flowcell (v2 chemistry). The sequence reads were aligned to the reference genome (hg18) with the Novoalign aligner. Duplicate reads, resulting from PCR clonality or optical duplicates, and reads mapping to multiple locations were excluded from downstream analysis. Single nucleotide substitutions and small insertion deletions were identified and quality filtered within the SamTools software package.20 Filtering of variants for novelty was performed by comparison to dbSNP131, 1000 Genomes SNP calls (March 2010).

PCR and Sanger sequencing

Primers were designed to cover exonic and intronic flanking sequences for NVL and GJC2 (primer sequences and PCR conditions are available upon request). PCR products were sequenced using BigDye Terminator v3.1 and an ABI 3130×1 Genetic Analyser. The sequencing traces were visually inspected in Finch TV v1.4 (Geospiza, Inc, Seattle, WA, USA) and compared to wild type reference sequence in CLC Sequence Viewer 6.4 (CLC bio A/S).
phenotype linked to this locus. Of mutations underlying the autosomal dominant lymphoedema individuals. These results strongly suggest that p.Met210Arg). This variant was not observed in 150 control occurrence of a non-synonymous variant (c.629T>T (p.S48L)) were also observed and shown to cosegregate in this family. While clinical family histories did not suggest that the two families were closely related, interrogation of the SNP genotypes generated for linkage mapping across the chr 1q42.1 linkage interval (eight SNPs) suggested the presence of a shared ancestral haplotype across this region in the two families (data not shown).

Further sequencing of NVL and GJC2 was undertaken in a cohort of 19 unrelated individuals with primary lymphoedema and revealed three further families with variants in GJC2 but no additional variants in NVL. The variants in GJC2 observed in this cohort comprised two proposita with the same Ser48Leu variant that was identified in families I and II, and a single occurrence of a non-synonymous variant (c.629T>G, p.Met210Arg). This variant was not observed in 150 control individuals. These results strongly suggest that GJC2 is the site of mutations underlying the autosomal dominant lymphoedema phenotype linked to this locus.

Clinical assessment of GJC2 mutation positive individuals
Family II
Four affected individuals from this family (figure 2), were clinically examined. They all had four limb lymphoedema, with one having additional involvement of the face and genitals. The most severely affected individual reported multiple episodes of cellulitis. The age of initial onset of oedema was variable but predominantly in childhood. In one individual, leg swelling developed in childhood but arm swelling was not noted until the age of 30 years. Lymph scans were performed in three individuals. All scans revealed that, although lymph channels could be clearly imaged, uptake of tracer in the lymph nodes at 2 h post-injection was notably impaired in all four limbs. Specifically, lymph transport over 2 h was substantially reduced. One member of the family had undergone venous duplex scans of the lower limbs. This revealed incompetent great saphenous veins bilaterally.

Family III
Two members of this family (father and daughter) were examined in clinic. A history of lymphoedema in several other family members was reported, consistent with autosomal dominant inheritance. Both members of this family who were examined had lower limb lymphoedema only. In the younger of the two family members the oedema significantly worsened at the age of 13 years, but there is a history of onset before this age, possibly at birth. Her father did not recognise he was affected until examined in clinic and therefore the true age of onset is not known. He had, however, suffered with varicose veins from a young age. Clinically the hands were not oedematous so only lower limb lymph scans were requested. These showed relatively normal main tract filling but with notably impaired uptake in the groin nodes at 2 h, consistent with distal hypoplasia of the lymphatics. Vein scans performed on both individuals revealed identical results, with pronounced incompetence of the great saphenous vein bilaterally.

Family IV
This Somalian family had two sisters affected by swelling to the knees bilaterally. Although their parents were distant cousins, their mother and maternal aunt were also reported to be affected—again consistent with autosomal dominant inheritance. The age of onset in both cases was 12 years. Results of four limb lymph scans and duplex vein scans are awaited.

Family V
This family showed clear autosomal dominant inheritance with four affected members in three generations. The age of onset was between the age of 4–40 years. Multiple episodes of cellulitis were a feature. Lymphoscintigraphy showed impaired drainage in all four limbs and a venous duplex scan revealed incompetence of the great saphenous vein bilaterally.

Details of the GJC2 mutations, and summary of the clinical details, are given in table 1.

Table 1 Clinical, lymphoscintigraphic, and venous parameters known for those patients with a mutation in GJC2

<table>
<thead>
<tr>
<th></th>
<th>Family I</th>
<th>Family II</th>
<th>Family III</th>
<th>Family IV</th>
<th>Family V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of affected</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Age of onset (years)</td>
<td>8–14</td>
<td>Birth to 10</td>
<td>0 to adult</td>
<td>Bilateral lower limb</td>
<td>Bilateral lower limb</td>
</tr>
<tr>
<td>Limbs affected</td>
<td>Bilateral lower limb, 4 limb</td>
<td>4 limb, one with face and genital swelling</td>
<td>Bilateral lower limb</td>
<td>4 limb oedema</td>
<td></td>
</tr>
<tr>
<td>Varicose/incompetent veins</td>
<td>One individual</td>
<td>One individual</td>
<td>Father</td>
<td>Venous duplex awaited</td>
<td>Yes, incompetent vein</td>
</tr>
<tr>
<td>Other associated abnormalities</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Lymphoscintigraphy</td>
<td>Poor uptake at 2 h. Distal hypoplasia</td>
<td>Poor uptake at 2 h. Distal hypoplasia</td>
<td>Poor uptake at 2 h. Distal hypoplasia</td>
<td>Awaited</td>
<td>Poor uptake at 2 h. Distal hypoplasia</td>
</tr>
</tbody>
</table>

DISCUSSION

This study clearly demonstrates the power and efficiency of the combination of linkage analysis with whole exome sequencing in the identification of causative genes in Mendelian phenotypes. Moving from the successful identification of the co-segregating chromosome 1 interval containing 173 genes to the identification of the causative mutation would previously have been prohibitively costly and time consuming, beyond Sanger sequencing of a small number of potential candidate genes. Recent developments in high throughput sequencing have provided a novel solution to such challenges. The approach undertaken here enabled rapid and cost effective sequencing of the entire exome, but interrogation of the generated sequence data was restricted to genes located within the linkage interval. This initial sequencing led to a shortlist of two potential variants which could not be differentiated within family I, but ultimately this was resolved through the identification of additional mutations in GJC2 in a cohort of unrelated lymphoedema cases. It is also notable that computational prediction of potential functional consequences of the two variants using the SIFT algorithm was of no benefit in this scenario. The results predicted the Val746Met variant in NVL to be damaging whereas the Ser48Leu variant in GJC2 was proposed to be tolerated. The use of computational predictors of the functional consequences of variants with algorithms including SIFT and PolyPhen have been proposed as a method of filtering variants for potential pathogenicity in high throughput sequencing studies. However, the experiences in our study and those of Ng et al. suggest that the interpretation of computational predictions must be handled with appropriate caution. Ultimately the success of the linkage and second generation sequencing approach in this study is clear, with the identification of GJC2 as the site of mutation in primary lymphoedema. This finding confirms the investigations of Ferrell et al., who used the results from a differential gene expression study to identify candidate genes in lymphoedema on the basis of greater expression in lymphatic endothelial cells (PECs) compared to blood endothelial cells (BECs). GJC2 was highlighted by their approach as its expression levels were five times higher in LECs than BECs. Subsequent sequencing of this gene in a large cohort (160 probands) of lymphoedema patients revealed six pathogenic mutations within the coding region of GJC2.

From a clinical viewpoint, individuals identified with heterozygous mutations in GJC2 characteristically presented with four limb lymphoedema, although some affected individuals only exhibited bilateral lower limb oedema. The age of onset was variable, both between and within families, ranging from birth to 40 years of age. It is possible that those affected with only lower limb lymphoedema may develop upper limb swelling at a later date. The upper limbs are always less severely affected and need to be examined specifically for lymphoedema. In those individuals subjected to lymphoscintigraphy, functional rather than anatomical abnormalities have been demonstrated. Thus, lymph drainage routes as seen on images looked normal but quantification revealed unequivocally reduced lymph transport as measured by ilio-inguinal/axillary node uptake over 2 h. This was a consistent finding in all five lymphoscintigrams thus far undertaken. Furthermore, reduced axillary lymph node uptake and impaired upper limb transport was evident in one individual without obvious hand oedema. Therefore, phenotyping for this category of lymphoedema requires a careful history and examination of hand oedema, as well as upper limb lymphoscintigraphy with quantification of lymph transport. The mechanism for lymph drainage failure needs further investigation.

Most of the affected individuals (particularly the adults) suffered with truncal varicose veins or had evidence of reflux of the great saphenous vein on duplex ultrasound examination. A true venous phenotype needs confirmation, but the findings support the view that genetic forms of lower limb lymphoedema are often associated with venous valve failure, as has been demonstrated with VEGFR3 and FOXC2 mutations. Venous reflux in adults may contribute to lower limb but not upper limb oedema, because of increased fluid filtration from venous hypertension. Greater gravitational stress (dependency) through higher venous pressures may explain increased expression of lower limb lymphoedema, with vulnerable upper limb drainage relatively protected from dependency.

Mutations in GJC2 have previously been described in the literature as the cause of autosomal recessively inherited Pelizaeus–Merzbacher-like disease (PMLD), which is a rare hypomyelinating disorder of the central nervous system. Uhlenberg et al noted that their heterozygous individuals had no neurological symptoms. PMLD is a very rare condition, so it is possible that mild lymphoedema could have been present in the parents, but it seems improbable that this would have passed unnoticed in all heterozygotes. It is more likely that the mechanism of action is different in the two conditions. Examples of genes causing different diseases where there is dominant or recessive inheritance of mutations include ROR2 in brachydactyly and Robinow syndrome and BSCL2 in spastic paraplegia type 17 and generalised lipodystrophy type 2.

In both these cases there appears to be a loss of function in the recessive conditions and a gain of function, or interference with the wild type product when there is dominant inheritance.

There is not yet sufficient evidence on the functional effects of GJC2 mutations in lymphoedema to know whether the same scenario applies here vis a vis PMLD. It is of interest that there appear to be at least two mechanisms for the effect of mutations in PMLD—namely, a loss of hemichannel function, or in some cases a dysfunction. To complicate the picture further, a mutation identified by Ferrell et al as causative in lymphoedema, p.Gly149Ser, is also reported as being causative for PMLD. A mechanism to explain this is currently difficult, especially as this mutation is one of those proposed to be loss of function of hemichannels in PMLD rather than dysfunction.

We conclude that mutations in GJC2 appear to be a significant cause of autosomal dominantly inherited four limb and bilateral lower limb lymphoedema. Further work will be required to determine the molecular mechanisms by which GJC2 contributes to the formation and maintenance of the lymphatic system.

Acknowledgements We are grateful to all of our patients and family members who have contributed to this and related lymphoedema genetic studies. FC and KK were supported by the British Heart Foundation, PD by the British Skin Foundation. The resources of the Biomics Unit at SGUL were essential for the study.

Funding The authors acknowledge the financial support from 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.

Competing interests None declared.

Ethics approval This study was conducted with the approval of the Wandsworth Local Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

Hum Genet indicate that CCBE1 is mutated in recessively inherited generalised lymphatic
Rapid identification of mutations in GJC2 in primary lymphoedema using whole exome sequencing combined with linkage analysis with delineation of the phenotype

Pia Ostergaard, Michael A Simpson, Glen Brice, et al.

*J Med Genet* published online January 25, 2011
doi: 10.1136/jmg.2010.085563

Updated information and services can be found at:
http://jmg.bmj.com/content/early/2011/01/24/jmg.2010.085563.full.html

These include:

**References**
This article cites 33 articles, 4 of which can be accessed free at:
http://jmg.bmj.com/content/early/2011/01/24/jmg.2010.085563.full.html#ref-list-1

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/