

Journal of Medical Genetics

Targeted Gene Panel Sequencing in Children with Very Early Onset Inflammatory Bowel Disease – Evaluation and Prospective Analysis

Journal:	<i>Journal of Medical Genetics</i>
Manuscript ID:	jmedgenet-2014-102624
Article Type:	Original Article
Date Submitted by the Author:	03-Jul-2014
Complete List of Authors:	Kammermeier, Jochen; UCL - Institute of Child Health, Genetics and Genomic Medicine; Great Ormond Street Hospital for Children, Gastroenterology Drury, Suzanne; Great Ormond Street Hospital for Children, North East Thames Regional Genetics Service James, Chela; UCL - Institute of Child Health, Genetics and Genomic Medicine Dziubak, Robert; Great Ormond Street Hospital for Children, Gastroenterology Ocaña, Louise; UCL - Institute of Child Health, Genetics and Genomic Medicine Elawad, Mamoun; Great Ormond Street Hospital for Children, Gastroenterology Beales, Philip L.; UCL - Institute of Child Health, Genetics and Genomic Medicine; UCL Institute of Child Health, London, Molecular Medicine Unit Lench, Nicholas; Great Ormond Street Hospital for Children, North East Thames Regional Genetics Service Uhlig, Holm; University of Oxford, Translational Gastroenterology Unit Bacchelli, Chiara; UCL - Institute of Child Health, Genetics and Genomic Medicine Shah, Neil; Katholic University Leuven, Gastroenterology; Great Ormond Street Hospital for Children, Gastroenterology
Keywords:	Genetics, Gastroenterology, Genetic screening/counselling, Inflammatory bowel disease, Molecular genetics

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Targeted Gene Panel Sequencing in Children with Very Early Onset
Inflammatory Bowel Disease – Evaluation and Prospective Analysis**

Jochen Kammermeier ^{1,2}, Suzanne Drury ³, Chela T James ⁴, Robert Dziubak
¹, Louise Ocaka ⁴, Mamoun Elawad ¹, Philip Beales ², Nicholas Lench ³, Holm
H Uhlig ^{5,6}, Chiara Bacchelli ² and Neil Shah ^{1,7}

1 Department of Gastroenterology, Great Ormond Street Hospital for
Children,
London, UK

2 Experimental & Personalised Medicine Section, Genetics and Genomic
Medicine, UCL Institute of Child Health, London, UK

3 North East Thames Regional Genetics Service, Great Ormond Street
Hospital for Children, London, UK

4 Centre for Translational Omics-GOSgene, Genetics and Genomic Medicine,
UCL Institute of Child Health, London, UK

5 Translational Gastroenterology Unit, University of Oxford, UK

6 Department of Pediatrics, University of Oxford, UK

7 Katholic University Leuven, Belgium

Key words

Very early onset inflammatory bowel disease, next generation sequencing,
whole exome sequencing

Corresponding author

J. Kammermeier (j.kammermeier@ucl.ac.uk)

University College London (Institute of Child Health) and Great Ormond Street
Hospital, 30 Guilford Street, London WC1N 1EH, England, UK

Word count

2191

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Background Multiple monogenetic conditions with partially overlapping phenotypes can present with IBD-like intestinal inflammation. With novel genotype-specific therapies emerging, establishing a molecular diagnosis is increasingly important.

Design We have introduced targeted next-generation sequencing technology as a prospective screening tool in children with very early onset IBD (VEOIBD). We evaluated the coverage of 40 VEOIBD genes in two separate cohorts undergoing targeted gene panel sequencing (TGPS) (n=25) and whole exome sequencing (WES) (n=20).

Results Targeted gene panel sequencing (TGPS) revealed causative mutations in four genes (*IL10RA*, *EPCAM*, *TTC37* and *SKIV2L*) discovered unexpected phenotypes and directly influenced clinical decision-making by supporting as well as avoiding haematopoietic stem cell transplantation. TGPS resulted in significantly higher median coverage when compared with WES, fewer coverage deficiencies and improved variant detection across established VEOIBD genes.

Conclusions Excluding or confirming known VEOIBD genotypes should be considered early in the disease course in all cases of therapy-refractory VEOIBD as it can have a direct impact on patient management. To combine both described NGS technologies would compensate for the limitations of WES for disease specific application whilst offering the opportunity for novel gene discovery in the research setting.

SIGNIFICANCE OF THIS STUDY

What is already known on this subject?

- A diverse group of monogenic conditions can present as very early onset IBD (VEOIBD).
- In the research setting, whole exome sequencing (WES) has facilitated the discovery of new VEOIBD genes and novel phenotypes of known conditions.
- Over 40 VEOIBD genes have been identified so far and are not routinely sequenced in patients with early onset and therapy-refractory disease.

What are the new findings?

- Targeted Gene Panel Screening (TGPS) revealed rare and unpredicted phenotypic variations.
- TGPS is a reliable genetic screening tool, leading to consistent coverage and variant detection across VEOIBD genes.
- In the diagnostic setting, WES performed with some limitations resulting in coverage deficiencies in several VEOIBD genes.

How might it impact on clinical practice in the foreseeable future?

- Comprehensive genetic screening in the diagnostic setting will reveal unexpected phenotypes, expand disease characterisation and open up new avenues for disease-specific treatments.
- In the advent of novel therapies for VEOIBD such as allogeneic haematopoietic stem cell transplantation (HSCT), confirming or

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

excluding known VEOIBD genotypes reliably and timely will be an essential requirement.

Confidential: For Review Only

INTRODUCTION

A diverse group of monogenic conditions can present as very early onset IBD (VEOIBD) ¹. Early molecular differentiation of VEOIBD has become increasingly important to determine the correct treatment pathway such as early hematopoietic stem cell transplantation (HSCT) in patients with IL10 signalling defects ², XIAP ³ or FOXP3 deficiency ⁴. Considering the orphan nature ⁵ and phenotype heterogeneity of VEOIBD, detecting a causative gene is challenging and can be both time and resource consuming. As a consequence, next-generation sequencing (NGS) technologies such as whole exome sequencing (WES) have been increasingly applied in patients with VEOIBD. Exome-wide screening has been proven invaluable as it revealed novel genotypes ⁶ and atypical phenotypes of known monogenic conditions ⁷. However, despite all efforts to ensure good overall exome coverage at reasonable costs, sequencing deficiencies persist in some genomic areas. An alternative NGS technology to WES is targeted gene panel sequencing (TGPS). It aims to enhance the capture of a specific selection of genes thus providing consistent and reliable sequence coverage.

To deploy NGS technologies in patients with VEOIBD in addition to deep phenotyping has not yet become routine practise. We therefore hypothesised that establishing a genotype timely in the disease course can have a significant impact on patient management.

In this study, we used TGPS as the first-line molecular diagnostic tool in children with VEOIBD and discuss the established genetic diagnoses and their clinical implications. Furthermore, we evaluated the sequencing accuracy from TGPS and WES for 40 VEOIBD genes and highlight specific technology

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

related weaknesses, which should be taken into consideration when applying
either technology in the diagnostic setting.

Confidential: For Review Only

MATERIAL AND METHODS

Patient cohort

Over a 12 month-period, 25 children were prospectively recruited for TGPS.

We included three patients with previously established genetic diagnoses (patient 5-7, table 1), which were all confirmed by TGPS (data not shown). All patients had extensive disease (rate of pancolitis or panenteritis: 100%) and were diagnosed within the first 36 months of life (median of 7 months [1,19]). Eighty per cent required long-term treatment with two or more immunosuppressant agents (table 1).

Patient	Ethnicity	CON	FHx	Age of Onset [months]	Disease Distribution	IS Agents	TPN	Surgery	TGPS
1	C	N	N	0	SB, P	2	Y	N	EPCAM
2	A	N	N	1	SB, P, PA	2	Y	N	IL10RA
3	A	Y	N	1	SB, P	3	Y	N	TTC37
4	A	Y	N	1	SB, P	5	Y	N	SKIV2L
5	A	Y	VEOIBD	4	SB, P, PA	2	Y	N	IL10RA*
6	A	Y	N	2	SB, P, PA	2	Y	Y	IL10*
7	C	N	N	36	SB, P	5	Y	N	XIAP*
8	A	Y	N	1	SB, P	2	N	Y	negative
9	C	N	N	20	P	1	N	N	negative
10	C	N	N	3	SB, P	2	Y	N	negative
11	C	N	N	12	SB, P	2	N	N	negative
12	C	N	N	7	SB, P	1	N	N	negative
13	C	N	N	12	SB, P	1	N	N	negative
14	C	N	N	0	SB, P	1	Y	Y	negative
15	A	N	N	36	SB, P	4	N	N	negative
16	A	Y	N	0	P	2	N	Y	negative
17	C	N	IBD	18	P	5	N	Y	negative
18	A	N	N	18	SB, P	2	N	N	negative
19	C	N	N	34	SB, P	5	N	N	negative
20	C	N	N	18	P	2	N	N	negative
21	A	N	N	3	SB, P	2	N	N	negative
22	C	N	N	15	SB, P	2	N	N	negative
23	A	Y	VEOIBD	0	P	2	N	N	negative
24	C	N	N	20	SB, P	1	N	N	negative
25	C	N	N	32	P	2	N	N	negative

Table 1. Phenotype characteristics for 25 patients analysed by TGPS. C: Caucasian, A: Asian, CON: consanguineous, FHx: family history, SB: small bowel, P: pancolonic disease, PA: perianal disease, IS Agents: immunosuppressive agents (steroids, azathioprine, cyclosporine, tacrolimus, sirolimus, infliximab, adalimumab) required during course of the disease, TPN: total parenteral nutrition required, Surgery: stoma formation and/or hemi-/colectomy. (*) Positive controls.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Targeted gene panel design

Protein coding sequences of 40 genes were selected to design the targeted gene panel (table 2). For 36 genes a comprehensive CCDS (Consensus Coding Sequence)⁸ was available. An alternative target (marked as * in table 1) was designed for the remaining genes based on all reported protein coding exons of gene transcripts published in Ensembl (European Bioinformatics Institute and Wellcome Trust Sanger Institute, Cambridge).

For *PIK3R1* and *SLC37A4*, the most comprehensive Ensembl gene transcripts ENST00000396611 and ENST00000357590 were selected. For *DCLRE1C* and *NCF4*, a target containing multiple transcripts was designed (*DCLRE1C*: CCDS31149 and first exon of ENST00000378289; *NCF4*: CCDS13935 and last exon of ENST00000248899 and first exon of ENST00000415063). Exon 1 of the *IKBKG* gene has not been captured in TGPS. This reflects the difference of one exon in two CCDS Sequences, reported for *IKBKG* (CCDS14757 and CCDS48196).

Exons are illustrated in numerical order following their strand orientation (+/-).

Group	Condition	Gene	CCDS
Epithelial Barrier Dysfunction	XL ectodermal dysplasia and immunodeficiency	<i>IKBKG</i>	CCDS48196
	ADAM-17-Deficiency	<i>ADAM17</i>	CCDS1665
	Familial Diarrhoea	<i>GUCY2C</i>	CCDS8664
	Tufting Enteropathy	<i>EPCAM</i>	CCDS1833
Phagocyte Defects	Chronic Granulomatous Disease	<i>CYBA</i>	CCDS32504
	Chronic Granulomatous Disease	<i>CYBB</i>	CCDS14242
	Chronic Granulomatous Disease	<i>NCF1</i>	CCDS34657
	Chronic Granulomatous Disease	<i>NCF2</i>	CCDS1356
	Chronic Granulomatous Disease	<i>NCF4</i>	NCF4*
	Glycogen Storage Disease Type 1b	<i>SLC37A4</i>	SLC37A4*
	Leukocyte Adhesion Deficiency Type 1	<i>ITGB2</i>	CCDS13716
Hyper/Auto-Inflammation	Phospholipase Cy2-Deficiency	<i>PLCG2</i>	CCDS42204
	Familial Hemophagocytic Lymphohistiocytosis Type 5	<i>STXBP2</i>	CCDS12181
	X-linked Lymphoproliferative Syndrome Type 2	<i>XIAP</i>	CCDS14606
	X-linked Lymphoproliferative Syndrome Type 1	<i>SH2D1A</i>	CCDS14608
	Hermansky-Pudlak Syndrome	<i>HPS1</i>	CCDS7475
	Hermansky-Pudlak Syndrome	<i>HPS4</i>	CCDS13835
	Hermansky-Pudlak Syndrome	<i>HPS6</i>	CCDS7527
T and B cell Defects	Common Variable Immunodeficiency Type 1	<i>ICOS</i>	CCDS2363
	Common Variable Immunodeficiency Type 8	<i>LRBA</i>	CCDS3773
	Agammaglobulinaemia	<i>BTK</i>	CCDS14482
	Agammaglobulinaemia	<i>PIK3R1</i>	PIK3R1*
	Hyper-IgM Syndrome	<i>CD40LG</i>	CCDS14659
	Hyper-IgM Syndrome	<i>AICDA</i>	CCDS41747
	Wiskott-Aldrich Syndrome	<i>WAS</i>	CCDS14303
	SCID/atypical SCID	<i>IL2RG</i>	CCDS14406
	SCID/atypical SCID	<i>RAG2</i>	CCDS7903
	SCID/atypical SCID	<i>LIG4</i>	CCDS9508
	SCID/atypical SCID	<i>CD3γ</i>	CCDS8395
	SCID/atypical SCID	<i>ZAP70</i>	CCDS33254
	SCID/atypical SCID	<i>ADA</i>	CCDS13335
	Omenn Syndrome	<i>DCLRE1C</i>	DCLRE1C*
	Hyper IgE syndrome	<i>DOCK8</i>	CCDS6440
Immune-Regulation	X-linked Immune Dysregulation, Polyendocrinopathy, Enteropathy (IPEX) Syndrome	<i>FOXP3</i>	CCDS14323
	IL10 Pathway Defects	<i>IL10</i>	CCDS1467
	IL10 Pathway Defects	<i>IL10RA</i>	CCDS8388
	IL10 Pathway Defects	<i>IL10RB</i>	CCDS13623
Others	Tricho-Hepato-Enteric Syndrome	<i>TTC37</i>	CCDS4072
	Tricho-Hepato-Enteric Syndrome	<i>SKIV2L</i>	CCDS4731
	Hirschprung's Disease	<i>RET</i>	CCDS7200

Table 2. 40 genes with VEOIBD phenotype ¹. Consensus code for each gene is shown (CCDS number). (*): No or insufficient CCDS available.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Next generation sequencing

20 samples were exome sequenced using a commercial service (Beckman Coulter Genomics), using SureSelect Human All Exon Kit Version 4 (Agilent Technologies). Samples were processed according to the SureSelect target Enrichment System for Illumina (see providers protocol) and run on an Illumina HiSeq2000 sequencer. All data sets fulfilled the company’s standardised quality criteria. 25 TGPS samples were sequenced in house using the SureSelect XT Custom Capture protocol (Agilent Technologies) on an Illumina MiSeq. Targeted capture baits for 40 VEOIBD genes were designed using the eArray tool (Agilent Biosystems, USA). Baits were designed for 162kb of sequence, with the use of 5x tiling and with a boosted number of baits for hard-to-capture, GC-rich regions.

NGS analysis pipeline

The samples were aligned to the human reference genome GRCh37/Hg19 with Burrows-Wheeler Aligner (BWA) software ⁹. The alignments were refined (base quality score recalibration, insertion-deletion realignment, duplicate removal) using the Genome Analysis Tool Kit (GATK) suite according to best practices suggested by the Broad Institute. Finally variants were called using UnifiedGenotyper (GATK) ¹⁰.

We applied a 10-read cut-off to account for the increased risk of unreliable heterozygous variant calling in compound heterozygous or autosomal dominant disease models at low depth. We defined deficient exons as those containing one or more base pairs covered by less than ten reads.

Three *in silico* prediction tools were used for predicting the damaging effect of the reported mutations (SIFT ¹¹, PolyPhen-2 ¹² and MutationTaster ¹³).

Sanger Sequencing

DNA for Sanger sequencing was extracted using the Chemagic-STAR (Hamilton, USA) and PCR performed with MegaMix (Microzone, UK) (custom primers sequences available on request; Sigma-Aldrich, UK). PCR products were purified using AmpureXP (Beckman Coulter Inc., UK) and sequenced with BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) followed by CleanSeq (Beckman Coulter Inc., UK). Sequences were established on the ABI3730XL (Applied Biosystems, USA). Traces were aligned to the reference sequence and variants called using Mutation Surveyor (SoftGenetics, USA).

Statistics

Statistical analysis was performed with IBM SPSS Statistics for Windows Version 22 (Armonk, NY). Categorical variables are expressed as proportions or frequencies and continuous variables as medians and interquartile ranges. Mann-Whitney U test was used to compare medians. For comparison of proportions, Pearson Chi-Square or Fisher's Exact Test was used where appropriate. All tests were two-tailed and the significance level was set at $p=0.05$.

Ethical approval

Patients were informed and consented for NGS as part of the "PETIT Study". The study had ethical approval from the National Research Ethics Service Committee London, Bloomsbury.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

Causative mutations established by TGPS

Using TGPS in a prospective setting, likely causative mutations were detected in four out of 22 patients (table 3, sanger sequencing traces: supplement 1). TGPS of patient 1 revealed a compound heterozygous mutation in the *EPCAM* gene. Both mutations (missense: p.Cys135Phe and deletion/frame shift: p.Thr234Lysfs*2) were not previously reported and were predicted damaging. Immunohistochemistry revealed the absence of EPCAM in intestinal epithelial cells and electron microscopy confirmed the diagnosis of tufting enteropathy (data not shown). Patient 2 harboured a novel and predicted damaging homozygous mutation in the *IL10RA* gene (p.Cys223Arg). Sequencing of patient 3 revealed compound heterozygosity for two mutations in the *TTC37* gene. The first variant, a predicted damaging heterozygous missense mutation has not been previously reported (p.Gly673Asp). The second variant resulting in the introduction of a premature stop codon (p.Trp936*), has been published in the homozygous state in a patient with tricho-hepato-enteric syndrome (THES) ¹⁴. Patient 4 harboured a homozygous splice site mutation in *SKIV2L*, a gene also associated with THES. *In silico* splice site mutation prediction with Alamut (Version 2.3) predicted likely skipping of exon 5 (c.355-2A>C).

Patient	Symptom Onset	CON	Clinical Features	Immune work-up	Histology	Treatment	Diagnosis pre-TGPS	Gene	Alleles	Mutation Analysis	Impact of Genotype
1 (male)	1 month	No	feeding intolerance intractable diarrhoea growth failure	normal	panenteric chronic inflammation villus blunting epithelial cell irregularity	steroids sirolimus PN	therapy-refractive panenteritis	<i>EPCAM</i>	c.404G>T (p.Cys135Phe) c.701del (p.Thr234Lysfs*2)	missense, ISP: damaging deletion/fs, ISP: damaging	HSCT abandoned IT considered genetic counselling
2 (male)	1 month	No	bloody diarrhoea perianal excoriation growth failure	normal	panenteric chronic inflammation most severe in the rectum	steroids PN	enterocolitis of infancy	<i>IL10RA</i>	c.667T>C (p.Cys223Arg) c.667T>C (p.Cys223Arg)	missense, ISP: damaging missense, ISP: damaging	HSCT genetic counselling
3 (male)	1 month	Yes	IUGR and prematurity intractable diarrhoea growth failure mild facial dysmorphism hepatic impairment sparse, fragile hair (no trichorrhexis nodosa)	normal	panenteric chronic inflammation villus atrophy crypt epithelial apoptosis	steroids azathioprine TNFα blockade PN	therapy-refractive panenteritis	<i>TTC37</i>	c.2018G>A (p.Gly673Asp) c.2808G>A (p.Trp936*)	missense, ISP: damaging stop gain, IPS: damaging	genetic counselling
4 (female)	1 month	Yes	IUGR and prematurity intractable diarrhoea feeding intolerance hypotonia and growth failure recurrent bacterial infections bronchiectasis hepatic impairment wiry hair (no trichorrhexis nodosa)	normal	panenteric chronic inflammation villus blunting epithelial cell irregularity	steroids azathioprine tacrolimus sirolimus TNFα blockade PN	therapy-refractive panenteritis	<i>SKIV2L</i>	c.355-2A>C c.355-2A>C	splice site prediction: skip of exon 5 ISP: damaging	genetic counselling

Table 3: Characteristics of patients with predicted pathogenic mutations on TGPS. CON: consanguineous, ISP: *in silico* prediction, PN: parenteral nutrition, IT: intestinal transplantation, "immune work-up normal": Immunoglobulin G/A/M/E levels, total neutrophil/lymphocyte count, lymphocyte subsets and neutrophil oxidative burst assay (NBT) all within normal range.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Comparison of base pair (bp) coverage: TGPS vs. WES

The median coverage within the protein coding regions of the 40 VEOIBD genes was significantly higher in TGPS samples when compared with WES (271.9 vs. 65.6, $p<0.001$ - figure 1). WES resulted in a higher rate of deficient exons (WES: 13.8% (1,594 / 11,520) vs. TGPS: 1.3% (182 / 14,400), $p<0.001$) as well as a higher rate of exons with complete absence of coverage (WES: 1.9% (215 / 11,520) vs. TGPS: 0.4% (52/14,400), $p<0.001$).

(FIGURE 1)

To evaluate how this trend reflects on individual gene level, we analysed the coverage data of all VEOIBD exons (supplement 2) and highlighted the results of six established VEOIBD genes (*FOXP3*, *IL10*, *IL10RA*, *IL10RB*, *XIAP* and *IKBKG*) in figure 2. For all six genes, the rate of coverage deficient exons was significantly higher in WES when compared to TGPS.

(FIGURE 2)

Consequence of coverage instability on variant detection

As a surrogate for bp-position specific reliability and sensitivity we compared common single nucleotide polymorphisms (SNP's) in three samples analysed by TGPS and WES. Eighty-nine SNP's were present in the protein coding target regions of the six samples. All polymorphisms were present in the TGPS samples with the corresponding WES samples failing to confirm six SNP's (in gene *ADA*, *CYBA*, *HPS1* and *NCF1*) (table 4).

Gene	Exon	SNP	TGPS			WES		
			Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
ADA	1	rs11555565	het	-	-	x	-	-
	2	rs394105	hom	hom	hom	hom	hom	hom
	4	rs11555566	-	-	het	-	-	het
	5	rs61737144	-	-	het	-	-	het
	6	rs244076	-	-	hom	-	-	hom
CYBA	1	rs8053867	hom	hom	hom	hom	x	hom
	4	rs4673	-	hom	-	-	hom	-
	6	rs1049254	hom	hom	hom	x	x	hom
HPS1	3	rs11539873	het	-	-	het	-	-
	5	rs1801287	het	het	-	het	het	-
	13	rs2296434	het	-	-	x	-	-
	16	rs2296436	het	-	-	het	-	-
NCF1	3	rs17856077	hom	het	het	x	x	x
	3	rs13447	-	het	het	-	x	x

Table 4. WES failed to call six SNP's (marked as "x") in four different genes. For each case (1-3) TGPS and corresponding WES SNP's (hom - homozygous, het - heterozygous) are shown.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

Children with VEOIBD rarely present with additional disease specific features such as early onset diabetes mellitus as observed in IPEX syndrome ¹⁵ or severe perianal disease in IL10 signalling defects ^{16 17}. Selective genetic analysis has become common practice in the presence of such features.

Over 40 identified VEOIBD genes and partly overlapping disease phenotypes ¹ render gene-by-gene sequencing futile and hence require a wider screening approach in order to establish a genetic diagnosis.

TGPS has already been successfully introduced in the genetic characterisation of conditions such as primary immune deficiencies ¹⁸. Our results suggest that TGPS is also an accurate genetic screening tool for children with VEOIBD. TGPS led to consistent coverage and variant detection across VEOIBD genes.

We could show that screening patients for established monogenic VEOIBD diseases reveals rare phenotypic variations: Congenital tufting enteropathy (CTE) has been described as an epithelial disorder without evidence of intestinal inflammation ¹⁹. As reported previously ²⁰, our results confirm that the presence of chronic inflammatory cells within the lamina propria does not exclude the diagnosis of CTE. Similarly, causative mutations in *SKIV2L* and *TTC37*, genes associated with tricho-hepato-enteric syndrome (THES), might lack key phenotypic features previously associated with the syndrome ^{14 21} suggesting that THES might be one of many possible phenotypes on the spectrum of variable gene penetrance involving the *SKIV2L/TTC37* genes.

1
2
3 In the advent of new therapeutic strategies such as HSCT for therapy-
4 refractory VEOIBD, establishing a genetic diagnosis early and accurately to
5 select potentially transplantable monogenic conditions, such as IL10 signalling
6 defects ², XIAP ³ or FOXP3 deficiency ⁴, is crucial. It is equally important to
7 exclude patients who are unlikely to benefit from such potentially life-
8 threatening therapy. In children with severe disease who have exhausted all
9 conventional treatment options HSCT has been considered despite the
10 absence of a genetic diagnosis ²². Even in these cases ruling out established
11 monogenic VEOIBD conditions such as XIAP deficiency helps the clinician to
12 decide on specific aspects of HSCT conditioning ³.

13
14 WES studies have revealed novel ⁶ and unexpected genotypes ⁷ and
15 highlighted new pathways involved in the pathophysiology of VEOIBD.
16 Designed for optimal overall exome coverage ²³, specific considerations have
17 to be made when WES is applied to investigate VEOIBD. Our study highlights
18 that WES might have some limitations in the diagnostic setting due to
19 coverage deficiencies in several VEOIBD genes. This is best exemplified by
20 the coverage of *IKBKG* (NEMO Deficiency) and *NCF1* (Chronic
21 Granulomatous Disease): *IKBKG* and *NCF1* baits have been frequently
22 omitted in commercially available WES captures to avoid nonspecific
23 alignment of reads to their pseudo gene loci leading to extensive coverage
24 dropout ^{24 25}. As a consequence, WES failed to detect all five SNP's in *NCF1*
25 established by TGPS.

26
27 Another reported phenomenon potentially leading to false negative results, is
28 the poor coverage of first exons ²⁶ in WES data, which we also observed for
29 VEOIBD genes.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

WES projects have revealed a number of likely causative mutations in novel VEOIBD genes but causality and clinical implication of such variants often remain to be established in lengthy follow-on studies. Genetic screening beyond the group of established VEOIBD genes for diagnostic purposes might therefore be less likely to have an immediate impact on patient care. Moreover, TGPS data requires less analysis time and storage capacity and excludes the likelihood of incidental findings and their associated ethical dilemmas²⁷.

When considering patients for genetic screening, conventional single-gene analysis might still be feasible in selected patients with distinct phenotypes. The nonspecific clinical profile of the majority of children however would favour NGS. In these cases sequential Sanger sequencing of potential candidate genes has been shown to far exceed the expenditure of time and expenses estimated for NGS analysis⁷.

Our data suggests that early comprehensive genetic screening can have a significant impact on patient management. Excluding or confirming known VEOIBD genotypes should therefore be considered early in the disease course in all cases of therapy-refractory VEOIBD.

Novel genetic platforms will facilitate to combine both NGS technologies, which would compensate for the limitations of WES for disease specific application whilst offering the opportunity for novel gene discovery in the research setting.

FIGURES AND SUPPLEMENTS

Figure 1: Coverage (TGPS vs. WES) within the protein coding regions of 40 VEOIBD genes.

Figure 2A: Six VEOIBD genes were compared individually. Black segments: exon contains complete loss of bp-coverage (0 reads), dark-grey segments: exon contains deficiency of bp-coverage (1-9 reads), light-grey segments: each bp of the exon is adequately covered (≥ 10 reads). Numbers within exon segments represent the samples affected by coverage deficiency/loss (20 WES and 25 TGPS samples). **2B:** For each gene, the percentage of deficient exons (WES vs. TGPS) and their rate of remaining adequate coverage (% of bp's covered by ≥ 10 reads) are shown.

Supplement 1. Sanger sequencing traces for the four described patients (1-4). For each mutation: top trace - reference, middle trace - forward strand, lower trace - reverse strand (aligned using Mutation Surveyor). Variants outlined in red.

Supplement 2. Comparison of coverage: WES (A) vs. TGPS (B).

Gene names as per HUGO nomenclature (HGNC), (*) indicates genes with custom-made alternative consensus. Red: exon contains complete loss of bp-coverage (0 reads), orange: exon contains deficiency of bp-coverage (1-9 reads), green: each bp of the exon is sufficiently covered (≥ 10 reads). Numbers within coloured exon segments represent the samples affected by coverage deficiency/loss (out of 20 WES and 25 TGPS patients). Only genes with loss/deficiencies of coverage in either WES or TGPS are shown (36 out of 40). Beyond exon 30, only affected exons in either WES or TGPS are presented (total number of exons > 30 in the following genes: *DOCK8*: 48 exons, *LRBA*: 57 exons, *PLCG2*: 32 exons and *TTC37*: 40 exons).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

AUTHORS' CONTRIBUTION

JK and CB were responsible for analysing the WES data, SD performed and analysed the TGPS data. CJ did the bioinformatics on the TGPS and WES raw data. RD performed statistical analysis. HHU and NS advised on the design of the study and the TGPS platform. JK drafted the first version of the manuscript with subsequent critical appraisal from all listed authors.

GRANT SUPPORT

JK is supported by a Great Ormond Street Charity Grant (GOSHCC: V1204). JK, BC, CJ, LO and PB are supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at Great Ormond Street Hospital for Children NHS Foundation Trust (GOSH) and University College London (UCL).

DISCLOSURE AND CONFLICT OF INTEREST

JK, SD, CB, TCJ, RD, ME, LO, HHU, NL, PB and NS declared no conflict of interest related to this article.

ACKNOWLEDGEMENTS

We thank the IBD and Clinical Genetics Teams at Great Ormond Street Hospital for their support in patient recruitment and sample management/preparation. This study was performed in partnership with GOSgene based at the UCL Institute of Child Health, and is supported by the National Institute for Health Research Biomedical Research Centre (NIHR BRC). This report is independent research by the NIHR BRC Funding Scheme. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

REFERENCES

1. Uhlig HH. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. *Gut* 2013;**62**(12):1795-805.
2. Engelhardt KR, Shah N, Faizura-Yeop I, Kocacik Uygün DF, Frede N, Muise AM, Shteyer E, Filiz S, Chee R, Elawad M, Hartmann B, Arkwright PD, Dvorak C, Klein C, Puck JM, Grimbacher B, Glocker EO. Clinical outcome in IL-10- and IL-10 receptor-deficient patients with or without hematopoietic stem cell transplantation. *The Journal of allergy and clinical immunology* 2013;**131**(3):825-30.
3. Marsh RA, Rao K, Satwani P, Lehmberg K, Muller I, Li D, Kim MO, Fischer A, Latour S, Sedlacek P, Barlogis V, Hamamoto K, Kanegane H, Milanovich S, Margolis DA, Dimmock D, Casper J, Douglas DN, Amrolia PJ, Veys P, Kumar AR, Jordan MB, Bleesing JJ, Filipovich AH. Allogeneic hematopoietic cell transplantation for XIAP deficiency: an international survey reveals poor outcomes. *Blood* 2013;**121**(6):877-83.
4. Barzaghi F, Passerini L, Bacchetta R. Immune dysregulation, polyendocrinopathy, enteropathy, x-linked syndrome: a paradigm of immunodeficiency with autoimmunity. *Frontiers in immunology* 2012;**3**:211.
5. Benchimol EI, Fortinsky KJ, Gozdyra P, Van den Heuvel M, Van Limbergen J, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflammatory bowel diseases* 2011;**17**(1):423-39.
6. Avitzur Y, Guo C, Mastropaolo LA, Bahrami E, Chen H, Zhao Z, Elkadri A, Dhillon S, Murchie R, Fattouh R, Huynh H, Walker JL, Wales PW, Cutz E, Kakuta Y, Dudley J, Kammermeier J, Powrie F, Shah N, Walz C, Nathrath M, Kotlarz D, Puchaka J, Krieger J, Racek T, Kirchner T, Walters TD, Brumell JH, Griffiths AM, Rezaei N, Rashtian P, Najafi M, Monajemzadeh M, Pelsue S, McGovern DP, Uhlig HH, Schadt E, Klein C, Snapper SB, Muise AM. Mutations in Tetratricopeptide Repeat Domain 7A Result in a Severe Form of Very Early Onset Inflammatory Bowel Disease. *Gastroenterology* 2014.
7. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genetics in medicine : official journal of the American College of Medical Genetics* 2011;**13**(3):255-62.
8. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ, Hart E, Suner MM, Landrum MJ, Aken B,

Ayling S, Baertsch R, Fernandez-Banet J, Cherry JL, Curwen V, Dicuccio M, Kellis M, Lee J, Lin MF, Schuster M, Shkeda A, Amid C, Brown G, Dukhanina O, Frankish A, Hart J, Maidak BL, Mudge J, Murphy MR, Murphy T, Rajan J, Rajput B, Riddick LD, Snow C, Steward C, Webb D, Weber JA, Wilming L, Wu W, Birney E, Haussler D, Hubbard T, Ostell J, Durbin R, Lipman D. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome research* 2009;**19**(7):1316-23.

9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**(14):1754-60.

10. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 2011;**43**(5):491-8.

11. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* 2003;**31**(13):3812-4.

12. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al]* 2013;**Chapter 7**:Unit7 20.

13. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* 2010;**7**(8):575-6.

14. Hartley JL, Zachos NC, Dawood B, Donowitz M, Forman J, Pollitt RJ, Morgan NV, Tee L, Gissen P, Kahr WH, Knisely AS, Watson S, Chitayat D, Booth IW, Protheroe S, Murphy S, de Vries E, Kelly DA, Maher ER. Mutations in TTC37 cause trichohepatoenteric syndrome (phenotypic diarrhea of infancy). *Gastroenterology* 2010;**138**(7):2388-98, 98 e1-2.

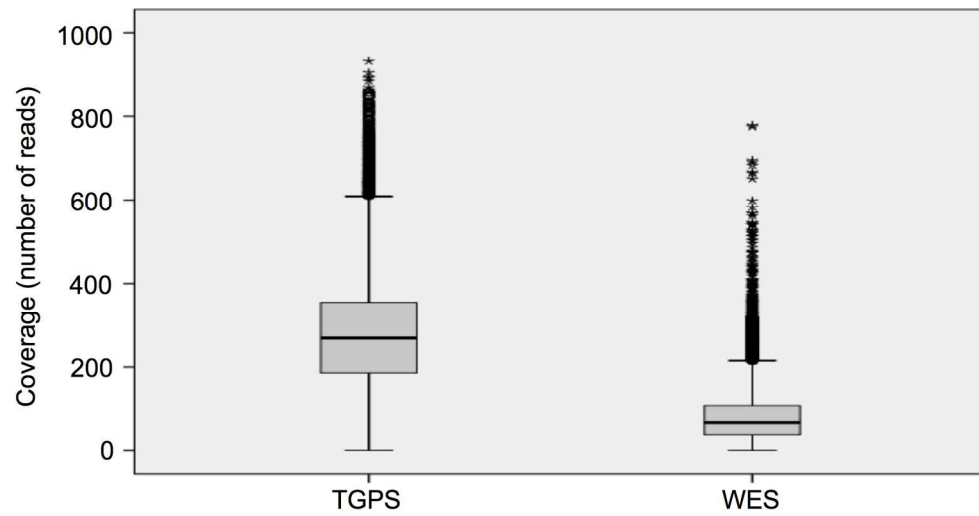
15. Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, Whitesell L, Kelly TE, Saulsbury FT, Chance PF, Ochs HD. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nature genetics* 2001;**27**(1):20-1.

16. Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schaffer AA, Noyan F, Perro M, Diestelhorst J, Allroth A, Murugan D, Hatscher N, Pfeifer D, Sykora KW, Sauer M, Kreipe H, Lacher M, Nustede R, Woellner C, Baumann U, Salzer U, Koletzko S, Shah N, Segal AW, Sauerbrey A, Buderus S, Snapper SB, Grimbacher B, Klein C. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *The New England journal of medicine* 2009;**361**(21):2033-45.

17. Glocker EO, Frede N, Perro M, Sebire N, Elawad M, Shah N, Grimbacher B. Infant colitis--it's in the genes. *Lancet* 2010;**376**(9748):1272.
18. Nijman IJ, van Montfrans JM, Hoogstraat M, Boes ML, van de Corput L, Renner ED, van Zon P, van Lieshout S, Elferink MG, van der Burg M, Vermont CL, van der Zwaag B, Janson E, Cuppen E, Ploos van Amstel JK, van Gijn ME. Targeted next-generation sequencing: A novel diagnostic tool for primary immunodeficiencies. *The Journal of allergy and clinical immunology* 2013.
19. Sivagnanam M, Mueller JL, Lee H, Chen Z, Nelson SF, Turner D, Zlotkin SH, Pencharz PB, Ngan BY, Libiger O, Schork NJ, Lavine JE, Taylor S, Newbury RO, Kolodner RD, Hoffman HM. Identification of EpCAM as the gene for congenital tufting enteropathy. *Gastroenterology* 2008;**135**(2):429-37.
20. Gerada J. Mucosal Inflammation as a Component of Tufting Enteropathy. *Immunogastroenterology* 2013;**2**(1):62-67.
21. Fabre A, Charroux B, Martinez-Vinson C, Roquelaure B, Odul E, Sayar E, Smith H, Colomb V, Andre N, Hugot JP, Goulet O, Lacoste C, Sarles J, Royet J, Levy N, Badens C. SKIV2L mutations cause syndromic diarrhea, or trichohepatoenteric syndrome. *American journal of human genetics* 2012;**90**(4):689-92.
22. Austen J. J. Worth, Zohreh Nademi, Jochen Kammermeier, Su Bunn, Robert Chiesa, Andrew Cant, Sophie Hambleton, Neil Shah, Mary Slatter, Kanchan Rao, Andrew Gennery, Mamoun Elawad, Persis Amrolia, Paul Veys. Allogeneic Stem Cell Transplant Offers Cure for Intractable Childhood Enteropathy. *Biology of Blood and Marrow Transplantation* 2014;**20**(2):S89-S90.
23. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. Performance comparison of exome DNA sequencing technologies. *Nature biotechnology* 2011;**29**(10):908-14.
24. Aradhya S, Bardaro T, Galgoczy P, Yamagata T, Esposito T, Patlan H, Ciccodicola A, Munnich A, Kenwrick S, Platzer M, D'Urso M, Nelson DL. Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. *Human molecular genetics* 2001;**10**(22):2557-67.
25. Heyworth PG, Noack D, Cross AR. Identification of a novel NCF-1 (p47-phox) pseudogene not containing the signature GT deletion: significance for A47 degrees chronic granulomatous disease carrier detection. *Blood* 2002;**100**(5):1845-51.
26. Kalari KR, Casavant M, Bair TB, Keen HL, Comeron JM, Casavant TL, Scheetz TE. First exons and introns--a survey of GC content and gene structure in the human genome. *In silico biology* 2006;**6**(3):237-42.

27. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* 2013;**15**(7):565-74.

Confidential: For Review Only



166x87mm (300 x 300 DPI)

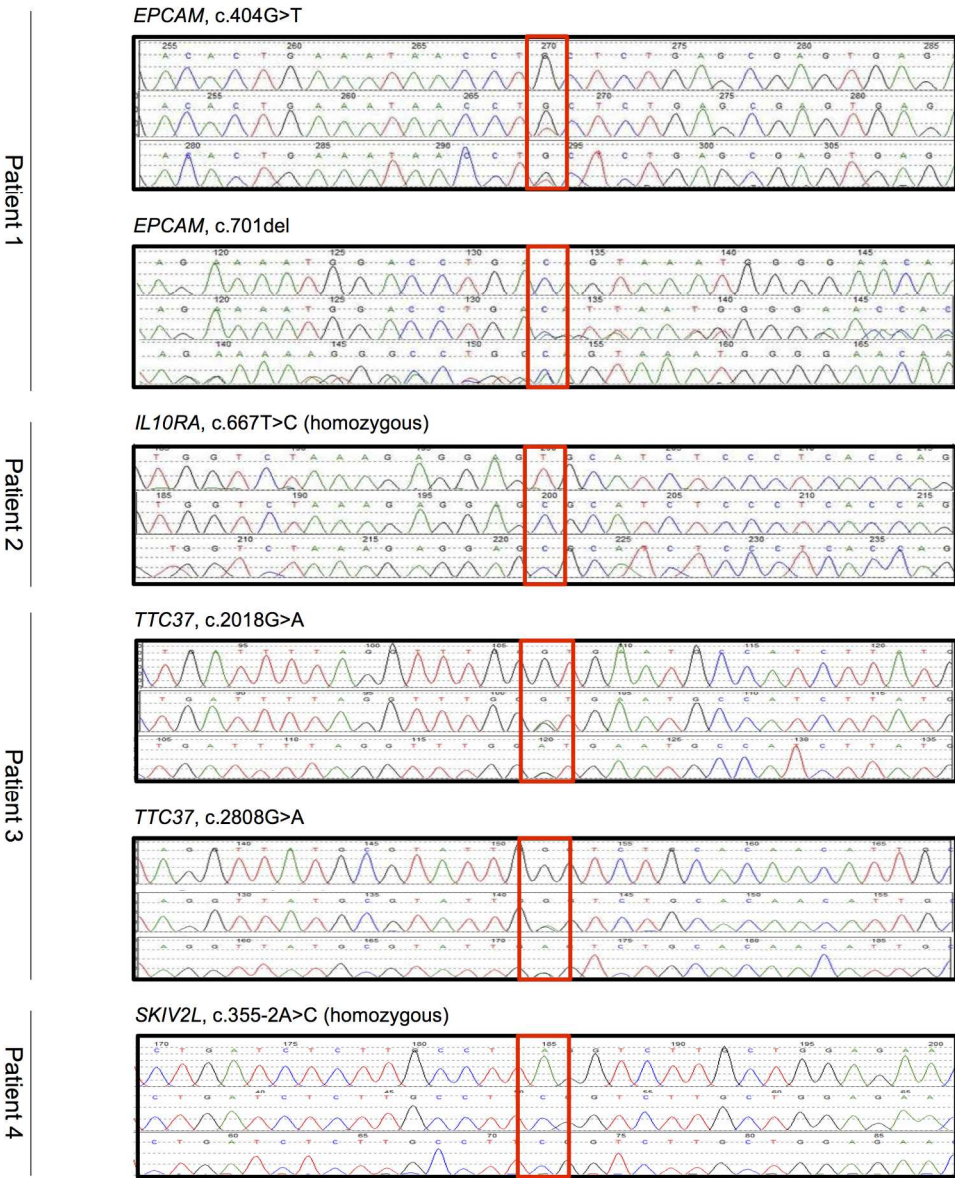
GENE	NGS Method	EXONS										
		01	02	03	04	05	06	7	08	09	10	11
FOXP3	WES	4 15	9	1 8	15			2	1 1			
	TGPS											
IKBKG	WES		4	14 6	18 2	16 4	13 7	14 6	16 4	18 2	15 5	
	TGPS	25				1						
IL10	WES				9							
	TGPS											
IL10RA	WES	4 3						1 2				
	TGPS											
IL10RB	WES	6 11		1		1						
	TGPS											
XIAP	WES	3		2 17	1	13		12 8				
	TGPS			4								

222x98mm (300 x 300 DPI)

GENE	% of deficient exons			% of adequate coverage within deficient exons [IQR]	
	WES	TGPS	p-value	WES	TGPS
<i>FOXP3</i>	25.5	0.0	<0.001	58.0 [16.7, 68.9]	N/a
<i>IKBKG</i>	82.0	10.4	<0.001	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
<i>IL10</i>	9.0	0.0	0.001	81.8 [75.0, 94.0]	N/a
<i>IL10RA</i>	7.1	0.0	<0.001	59.7 [38.4, 87.3]	N/a
<i>IL10RB</i>	13.6	0.0	<0.001	0.0 [0.0, 28.6]	N/a
<i>XIAP</i>	46.7	2.7	<0.001	0.0 [0.0, 81.8]	92.4 [34.8, 98.7]

246x89mm (300 x 300 DPI)

trial: For Review Only



176x213mm (300 x 300 DPI)

