Specifications of the NGS panel Hemoglobinopathy

Targeted sequencing with the Ion Torrent System is able to identify single nucleotide variants, small insertions and small deletions. Variants in repeat sequences, large homopolymers and large insertions/deletions are not or difficult to identify.

The RBC Hemoglobinopathy AmpliSeq Panel (IAD127644_241) exist of 141 primer pairs, which amplify 151 different amplicons. Five primer pairs were designed based on HBA1 gene coordinates and are also 100% specific for the HBA2 gene and five primer pairs were designed based on HBG1 gene coordinates and are also 100% specific for the HBG2 gene. The panel design is covering 35.28 Kbase, and includes almost all coding regions (exons), flanking intronic regions, untranslated regions and known promotor regions of the genes of interest as depicted in Table 1. Also genomic regions that are involved in the expression of the globins and in fetal γ - to adult β -globin switching, are included in the design, as listed in Table 1.

Table 1: List of genes and genomic regions and the coverage of these regions

Genes	Chromosome	NCBI Transcript	Exons	Coverage	Coverage 5' of initiation site (flanking sequence and 5'UTR)	Amplicons
HBA1	chr16	NM_000558.4	3	100%	134 bases before initation site (incl. promotor)	5
HBA2	chr16	NM_000517.4	3	100%	134 bases before initation site (incl. promotor)	6
HBB	chr11	NM_000518.4	3	100%	242 bases before initation site (incl. promotor)	10
HBD	chr11	NM_000519.3	3	100%	174 bases before initation site (incl. promotor)	7
HBG1	chr11	NM_000559.2	3	100%	458 bases before initation site (incl. promotor)	5
HBG2	chr11	NM_000184.2	3	100%	461 bases before initation site (incl. promotor)	5
BCL11A	chr2	NM_022893	4	100%	440 bases before initation site	20
KLF1	chr19	NM_006563.3	3	100%	202 bases before initation site	8
GATA1	chrX	NM_002049	5	100%	250 bases before initation site	11
ASF1B	chr19	NM_018154	4	100%	125 bases before initation site	4
E2F2	chr1	NM_004091	7	100%	18 bases before initation site	7
E2F4	chr16	NM_001950	10	100%	215 bases before initation site	11
MYB	chr6	NM_001130173.1	16	100%	128 bases before initation site	18
ZBTB7A	chr19	NM_015898	2	98,25%	94 bases before initation site	9
Regions	**Chromosome**	**GRCh37/hg19 coordinates**	**Regions**	**Coverage**		**Amplicons**
α-MRE HS-40	chr16	163462 - 163792	1	100%		4
LCRB-3'HS1	chr11	5226054 - 5226417	1	100%		3
LCRB-HS1	chr11	5297694 - 5297904	1	100%		1
LCRB-HS2	chr11	5302012 - 5302212	1	100%		1
LCRB-HS3	chr11	5305937 - 5306117	1	100%		1
LCRB-HS4	chr11	5309551 - 5309744	1	100%		1
LCRB-HS5	chr11	5312524 - 5312732	1	100%		1
LCRB-HS111	chr11	5402250 - 5401770	1	100%		4
BCL11A erythroid enhancer 1	chr2	60718035 - 60718305	1	100%		2
BCL11A erythroid enhancer 2	chr2	60722235 - 60722535	1	100%		3
BCL11A erythroid enhancer 3	chr2	60725235 - 60725735	1	99,80%		3
HBS1L-MYB	chr6	135431600 - 135431680	1	97,50%		1

For ZBTB7A gene, 31 bases are missed in exon 1 and 2, divided over 4 different amplicons. Also 2 bases and 1 bases of the regions MYB- HBS1L and BCL11A-enhancer3 are missed in the design. There are no disease causing variants published in these missing areas as there are no consequences reported for copy number variants (HGMD version 2017.2). See Table 2.

Table 2: The panel design does not cover the following bases

Genes	Coverage	Number Missing Bases	GRCh37/hg19 coordinates	Description	Missing HGMD-DM variants
ZBTB7A	98,25%	5	4047887 - 4047892	central in exon 2	non
		12	4054253 - 4054265	central in exon 1	non
		5	4054744 - 4054749	central in exon 1	non
		9	4055044 - 4055053	central in exon 1	non
BCL11A erythroid enhancer 3	99,80%	1	60725379 - 60725380	1 base missing in central area enhancer 3	non
HBS1L-MYB	97,50%	2	135431678 - 135431680	2 bases missing at 3'site	non

In Table 3 is listed the published disease causing variants HGMD version 2017.2) which are not covered in the designed panel. Including these regions in the design failed caused by high sequence homology between HBB and HBD gene.

Table 3: List of missed HGMD disease causing variants in the designed Hb panel

Gene	Chromosome	GRCh37/hg19	HGMD ref.	HGVS	Variant class	Phenotype	Reference
HBB	chr11	5248052	CS001426	c.93-23T>C	DM	Thalassaemia beta	Muniz (2000) Am J Hematol 64:7
HBB	chr11	5248050	CS810003	c.93-21G>A	DM	Thalassaemia beta	Spritz (1981) Proc Natl Acad Sci U S A 78:2455
HBD	chr11	5255567	CS109514	c.92+5G>T	DM	Thalassaemia delta	Amirian (2010) Hemoglobin 34:594

For accepting a sequence run with the Hemoglobinopathie AmpliSeq panel, the %Base20x of all amplicons must be at least **99.2%**. With this acceptance criteria, the first 60 bases of exon 1 and the complete 5'UTR of the HBA2 gene will be less covered than 20 times per base. This may result in low coverage for 35 DM and 3 DM? variants located in this region (see Table 4a&b). Also small regions in HBG1, GATA1 and BCL11A have a low coverage, but without missing published DM variants, as listed in Table 4a.

Preferably, a %Base20x coverage of at least **99.7%** must be reached to insure that also exon 1 of gene HBA2 is completely covered. Under these condition, ≥99.7% coverage, only 3DM and 1DM? variant located in the 5'UTR region will have a low coverage.

Table 4: Failed amplicons and their missing published mutations at %Base20x coverage of 99.2%

Gene	Amplicon	%Base20x amplicon	Locations with base coverage below 20x		missing HGMD
			chromosomal coordinates	description	
HBA2	ES7.HBA21_1	12%	chr16: 222775 - 222971	missing first 60 bases of exon 1 and complete UTR (65 bases)	35 DM and 3 DM?
HBG1	AMPL7160773269	46%	chr11: 5270577 - 5270689	missing last 100 bases of exon 2 and 12 bases of intron 2	none
GATA1	AMPL7164888082	22%	chrX: 48649168 - 48649403	missing 235 bases of intron 1, located 94 bases before UTR of exon2	none
BCL11A	ES7.BCL11A_14	50%	chr2: 60689648 - 60689749	missing 100 bases of intron 3, located 87 bases before start exon 4	none

Table 5: HGMD DM and DM? variants of gene HBA2 missed at %Base20x coverage of 99.2% and 99.7%

Coverage <20x %Base20x	GRCh37/hg19	HGMD Ref	HGVS	Variant class	Phenotype	Reference
99.2% / 99.7%	chr11:222821	CR140643	g.222821G>A	DM?	Microcytosis	Qadah (2014) Pathology 46:46
99.2% / 99.7%	chr11:222853	CR140642	g.222853C>T	DM	Haemoglobin variant	Qadah (2014) Pathology 46:46
99.2% / 99.7%	chr11:222889	CR1310302	g.222889C>G	DM	Thalassaemia alpha	Yao (2013) Gene 532:120
99.2% / 99.7%	chr11:222891	CR042845	g.222891C>G	DM	Haemoglobin variant	Lacerra (2004) Hum Mutat 24:338
99.2%	chr11:222910	CD032743	c.-3_-2delAC	DM	Thalassaemia alpha	Morle (1985) EMBO J 4:1245
99.2%	chr11:222910	CR051278	g.222910C>T	DM	Haemoglobin variant	Sarkar (2005) Br J Haematol 129:282
99.2%	chr11:222912	CM870032	c.1A>G	DM	Haemoglobin H disease	Olivieri (1987) Blood 70:729
99.2%	chr11:222912	CD066365	c.1delA	DM	Thalassaemia alpha	Eng (2006) Hemoglobin 30:149
99.2%	chr11:222913	CM840002	c.2T>C	DM	Thalassaemia alpha	Piratsu (1984) J Biol Chem 259:12315
99.2%	chr11:222913	CM076227	c.2T>G	DM	Haemoglobin variant	Hadavi (2007) Haematologica 92:992
99.2%	chr11:222913	CD973333	c.2delT	DM	Haemoglobin H disease	Waye (1997) Hemoglobin 21:469
99.2%	chr11:222914	CM1511050	c.3G>T	DM	Haemoglobin variant	de la Fuente-Gonzalo (2015) Clin Chem Lab Med epub
99.2%	chr11:222915	CM140705	c.4G>C	DM	Thalassaemia alpha	van Zwieten (2014) Hemoglobin 38:1
99.2%	chr11:222915	CM1412328	c.4G>A	DM?	Thalassaemia alpha	Alizadeh (2014) Clin Lab 60:941
99.2%	chr11:222916	CM021274	c.5T>C	DM	Haemoglobin variant	Lancan (2002) Am J Hematol 69:214
99.2%	chr11:222933	CM131529	c.22A>T	DM	Thalassaemia alpha	Bayat (2013) Hemoglobin 37:148
99.2%	chr11:222933	CM058058	c.22A>G	DM	Haemoglobin variant	Ngiwsara (2005) Hemoglobin 29:155
99.2%	chr11:222935	CM024472	c.24G>C	DM	Haemoglobin variant	Wajcman (1994) Hemoglobin 18:427
99.2%	chr11:222938	CD121675	c.27delC	DM	Thalassaemia alpha	Tang (2012) Hemoglobin 36:192
99.2%	chr11:222939	CM1310300	c.28A>T	DM	Thalassaemia alpha	Yao (2013) Gene 532:120
99.2%	chr11:222940	CM020945	c.29A>G	DM	Haemoglobin variant	Troxler (2002) Biochem Biophys Res Commun 292:1044
99.2%	chr11:222941	CM020944	c.30C>G	DM	Haemoglobin variant	Hoyer (2002) Hemoglobin 26:175
99.2%	chr11:222948	CM1515738	c.37G>C	DM	Anaemia, hypochromic microcytic	Farashi (2015) Hemoglobin 39:398
99.2%	chr11:222949	CM940901	c.38C>A	DM	Haemoglobin variant	Molchanova (1994) Br J Haematol 88:300
99.2%	chr11:222957	CD150815	c.41_46delCCTGGG	DM	Thalassaemia alpha intermedia	Kattamis (2015) Hemoglobin 39:55
99.2%	chr11:222954	CM830035	c.43T>C	DM	Haemoglobin variant	Moo-Penn (1983) Biochim Biophys Acta 747:65
99.2%	chr11:222957	CM1310301	c.46G>A	DM	Thalassaemia alpha	Yao (2013) Gene 532:120
99.2%	chr11:222960	CM940902	c.49A>G	DM	Haemoglobin variant	Molchanova (1994) Br J Haematol 88:300
99.2%	chr11:222962	CM820015	c.51G>T	DM	Haemoglobin variant	Liang (1982) Hemoglobin 6:629
99.2%	chr11:222963	CM169994	c.52G>T	DM	Thalassaemia alpha	Yang (2016) Hemoglobin 40:264
99.2%	chr11:222966	CM1212400	c.55G>C	DM	Haemoglobin variant	Griffiths (1977) FEBS Lett 75:93
99.2%	chr11:222967	CM1414884	c.56G>A	DM	Haemoglobinopathy, beta	Hassan (2014) Hemoglobin 38:422
99.2%	chr11:222967	CD102836	c.56delG	DM	Thalassaemia alpha	Al-Gazali (2010) Hum Mutat 31:505
99.2%	chr11:222971	CD033199	c.60delG	DM	Thalassaemia alpha	Harteveld (2003) Am J Hematol 74:99
99.2%	chr11:222972	CM004608	c.61C>G	DM	Haemoglobin variant	Prehu (2000) Hemoglobin 24:305
99.2%	chr11:222974	CM940903	c.63C>A	DM	Haemoglobin variant	Molchanova (1994) Br J Haematol 88:300
99.2%	chr11:222975	CM077578	c.64G>T	DM?	Thalassaemia alpha	Harteveld (2007) Hemoglobin 31:325
99.2%	chr11:222975	CM890288	c.64G>C	DM	Haemoglobin variant	Wajcman (1989) Hemoglobin 13:421

Reporting

Only clinical relevant variants will be reported. Variants with classification *Certainly Pathogenic* (class 5) and *Likely Pathogenic* (class 4) are always reported. Variants with category *Unknown significance* (class 3) will only be reported if the variant is expected to be involved in the phenotype of the patient. Category *Certainly Benign* (class 1) and *Likely Benign* (class 2) variants will not be reported. (see: *http://www.acgs.uk.com/quality/best-practice-guidelines/Variant Guidelines*. *Document: Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics*).

Besides reporting the clinical relevant variants we report whether a patients is heterozygous, homozygous, expected compound heterozygous or hemizygous for a mutation and how this may relate to disease phenotype.
All the variants are annotated and reported as designated by the Human Genome Variation Society (HGVS) nomenclature, as described at their website http://varnomen.hgvs.org