Immunogenetics (2017) 69:359–369
DOI 10.1007/s00251-017-0981-y

ORIGINAL ARTICLE

Mutational profiling in the peripheral blood leukocytes of patients with systemic mast cell activation syndrome using next-generation sequencing

Janine Altmüller[1,2] · Britta Haenisch[3] · Amit Kawalia[1] · Markus Menzen[4] · Markus M. Nöthen[5] · Heide Fier[6] · Gerhard J. Molderings[5]

Received: 27 February 2017 / Accepted: 26 March 2017 / Published online: 6 April 2017

Abstract Mast cell activation syndrome (MCAS) and systemic mastocytosis (SM) are two clinical systemic mast cell activation disease variants. Few studies to date have investigated the genetic basis of MCAS. The present study had two aims. First, to investigate whether peripheral blood leukocytes from MCAS patients also harbor somatic mutations in genes implicated in SM using next-generation sequencing (NGS) technology and a relatively large MCAS cohort. We also addressed the question, whether some of the previously as somatic reported mutations are indeed germline mutations. Second, to identify germline mutations of relevance to MCAS pathogenesis. Here, mutation frequency in the present MCAS cohort was compared to that in public- and in-house databases in the case of frequent variants, and co-segregation was investigated in multiply affected families in the case of rare variants (allele frequency < 1%). MCAS diagnoses were assigned according to current criteria. Twenty five candidate genes were selected on the basis of published findings for SM. NGS was performed using a 76kbp custom designed Agilent SureSelect Target Enrichment and an Illumina Hiseq2000 2x100bp sequencing run. NGS revealed 67 germline mutations. No somatic mutations were detected. None of the germline mutations showed unequivocal association with MCAS. Failure to detect somatic mutations was probably attributable to the dilution of mutated mast cell DNA in normal leukocyte DNA. The present exploratory association findings suggest that some of the detected germline mutations may be functionally relevant and explain familial aggregation. Independent replication studies are therefore warranted.

Keywords Mast cell activation disease · Mast cell activation syndrome · Systemic mastocytosis · Germline mutation · Next-generation sequencing

Electronic supplementary material The online version of this article (doi:10.1007/s00251-017-0981-y) contains supplementary material, which is available to authorized users.

✉ Gerhard J. Molderings
molderings@uni-bonn.de

[1] Cologne Center for Genomics, University of Cologne, D-50931 Cologne, Germany

[2] Center for Molecular Medicine Cologne, University of Cologne, D-50937 Cologne, Germany

[3] German Center for Neurodegenerative Diseases (DZNE), D-53127 Bonn, Germany

[4] Division of Internal Medicine, Department of Gastroenterology and Diabetology, Community Hospital, D-53113 Bonn, Germany

[5] Institute of Human Genetics, University Hospital of Bonn, D-53127 Bonn, Germany

[6] Institute of Genomic Mathematics and Bioinformatics, University of Bonn, D-53127 Bonn, Germany

Introduction

Mast cell activation syndrome (MCAS) and systemic mastocytosis (SM) belong to a group of primary mast cell diseases (Molderings et al. 2011; Hamilton et al. 2011; Valent et al. 2012) termed systemic mast cell activation disease (MCAD; Online Resource 1). Recent research suggests that the MCAD variants and clinical subtypes represent varying manifestations of a common process of mast cell dysfunction (Molderings et al. 2007, 2010; Hermine et al. 2008; Akin et al. 2010).

Previous studies of SM have identified several somatic KIT mutations, including the mutation KIT^{D816V} in isolated mast cells and purified bone marrow cell populations (Nagata et al. 1995; Longley et al., 1999; Garcia-Montero et al. 2006). Subsequently, these mutations have also been identified in

peripheral blood leukocytes from SM patients with an immature mast cell immunophenotype, (Kristensen et al. 2012; Teodosio et al. 2012). Recent next-generation sequencing (NGS) studies of SM have demonstrated somatic mutations in a large number of other genes. These genes encode proteins for various signaling pathways, epigenetic regulators, transcription factors, and the RNA splicing machinery (for review, see Molderings 2015).

In MCAS patients, studies of mast cell-committed progenitor cells isolated from peripheral blood also have identified several somatic KIT mutations (for review, see Molderings 2015). The identified mutations did not include KIT^{D816V}. In addition, one somatic mutation in *NLRP3* has been reported (for review, see Molderings 2015).

The observed familial aggregation of MCAD suggests that germline variants also contribute to disease development. In a previous systematic family study by our group, which included 84 MCAD patients, approximately 74% of the index patients had at least one first-degree relative with MCAD, irrespective of MCAD variant or gender (Molderings et al. 2013). The prevalence of MCAD among first-degree relatives was 46%, which differed significantly from the prevalence in the control group (approximately 17%; Molderings et al. 2013).

The present systematic investigation of the genetic background of MCAS had two aims. First, to determine whether peripheral blood leukocytes from MCAS patients harbor somatic mutations in genes implicated in SM. Here, NGS was used for the detection of mutations, since a low number of somatic mutations in peripheral blood leukocytes was assumed. Second, to identify germline mutations of relevance to MCAS pathogenesis. Germline and somatic mutations were distinguished by assessing the ratio of mutated to non-mutated sequences in the NGS step. The potential involvement in MCAS of the identified germline mutations was investigated by assessing: (i) allele frequency differences between the present MCAS patients and controls from public and in-house genetic variant databases in the case of frequent mutations; and (ii) co-segregation with disease in multiply affected families in the case of rare variants (allele frequency < 1% in the 1000 Genomes Project).

Methods

Patients and reference groups

The present cohort comprised 95 German MCAS patients (for details, see Table 1). These subjects presented to the *Bonn Interdisciplinary Research Group for Systemic Mast Cell Diseases* between May 2005 and December 2013 for diagnostic evaluation, and were assigned a diagnosis of a MCAS in accordance with current criteria (Molderings et al. 2011; Hamilton et al. 2011; Valent et al. 2007, 2012; Table 2). The age of the patients ranged from 16 to 85 years (mean: 47 years; male to female ratio: 1:1.8). The diagnostic criteria for MCAS and the clinical characteristics of the sample are listed in Table 2. As part of the diagnostic evaluation, the presence of mast cell mediator release syndrome was assessed by a a specialist in internal medicine using a validated questionnaire (Alfter et al. 2009; Molderings et al. 2013). Unrelated diseases with a similar presentation were excluded using appropriate assessments, including laboratory testing, imaging, and/or endoscopy. Patient data were pseudonomized prior to analysis. To enable evaluation of the familial co-segregation of detected variants, 14 affected relatives from seven multiply affected families were included. Two of these individuals were male monozygotic twins. Thus in total, 81 patients were unrelated.

Table 1 Characteristics of the study cohort

MCAS (n = 95)	
male (n = 34; 36%)	female (n = 61; 64%)
age [years]: mean ± SD, median, range	
41 ± 17, 41, 16–85	50 ± 15, 50, 19–75

MCAS – mast cell activation syndrome, SD - standard deviation

The control sample for the present study was drawn from the 1000 Genomes data, Phase 3 (1000 Genomes Project Consortium, 2015). For this purpose, the genotypic information of all 503 European sequenced individuals was extracted (Iberian, n = 107; Tuscanian, n = 107; British, n = 91; Finnish, n = 99; and residents of Utah with Northern and Western European ancestry, n = 99).

As an in-house reference group representative of the German general population, a sample of 1657 randomly recruited German subjects was used. These subjects underwent genetic evaluation in previous studies at the participating institutes using the NGS technique applied in the present patient group. Thus data comparability was considered optimal. However, due to the structure of the in-house data base, subject gender and mutations with a minor allele frequency of >8% were not available for the present analyses.

As a further reference group, the ExAC data were used. These data were obtained from a total of 60,706 individuals (http://exac.broadinstitute.org). The populations and cohorts represented in the ExAC data are listed at http://exac.broadinstitute.org/faq.

Mutation analysis

The following 25 candidate genes were selected on the basis of published findings for SM (Schwaab et al. 2013; Hanssens et al. 2014): *NRAS*, *KRAS*, *NLRP3*, *IDH1*, *IDH2*, *KIT*, *JAK2*, *CBL*, *DNMT3A*, *IL13*, *Il4*, *TNF*, *EZH2*, *MS4A2*, *KMT2A*,

Table 2 Percentage of the study population fulfilling the current proposed criteria for a diagnosis of mast cell activation syndrome (MCAS) (Afrin et al. 2016; for references, see text) following the exclusion of all differential diagnoses. A diagnosis MCAS was assigned upon fulfillment of the major criterion plus at least one minor criterion, or the fulfillment of at least three minor criteria. The number of patients fulfilling the respective criterion over the total number of patients included in the present study ($n = 95$) is indicated in parentheses

Proposed criteria for a diagnosis of mast cell activation syndrome		Percentage
Major criterion		
	Constellation of clinical symptoms attributable to a pathological increase in mast cell activity (mast cell mediator release syndrome)	100% (95/95)
Minor criteria		
1.	Focal or disseminated increase in the number of mast cells in marrow and/or extracutaneous organ(s) (e.g., gastrointestinal tract biopsies; CD117-, tryptase- and CD25-stained)	45% (43/95)
2.	Abnormal spindle-shaped morphology in >25% of mast cells in marrow or other extracutaneous organ(s)	5% (5/95)
3.	Abnormal mast cell expression of CD2 and/or CD25 (i.e., co-expression of CD117/CD25 or CD117/CD2)	5% (5/95)
4.	Detection of genetic changes in mast cells from blood, bone marrow, or extracutaneous organs for which an impact in terms of an increased activity in affected mast cells has been proven.*	18% (17/95)
5.	Evidence (typically from body fluids such as whole blood, serum, plasma, or urine) of raised levels of mast cell mediators including:	
	• tryptase in blood	21% (20/95)
	• histamine in blood or its metabolites (e.g., N-methylhistamine) in urine	23% (22/95)
	• heparin in blood	66% (63/95)
	• chromogranin A in blood (following the exclusion of potential confounders, i.e., cardiac or renal failure, neuroendocrine tumors, or recent proton pump inhibitor use)	17% (16/95)
	• other relatively mast-cell-specific mediators (e.g., eicosanoids including prostaglandin PGD_2 and its metabolite 11-β-$PGF_{2\alpha}$, or leukotriene E4)	28% (27/95)
6.	Symptomatic response to inhibitors of mast cell activation or mast cell mediator production or action (e.g., histamine H_1 and/or H_2 receptor antagonists, cromolyn)	84% (80/95)

*The present genetic findings were not used to assign a diagnosis of MCAS

U2AF1, *SRSF2*, *SF3B1*, *SETBP1*, *RASGRP4*, *TP53*, *TET2*, *ETV6*, *RUNX1,* and, *ASXL1*. NGS was performed using a 76kbp custom designed Agilent SureSelect Target Enrichment system (Agilent Technologies, Sante Clara, United States), and an Illumina Hiseq2000 2x100bp sequencing run (Illumina, San Diego, United States). The average data output per sample was 350 Mb, and an average on-target coverage of 1386× (range 825× − 1725×) without dropouts (all nucleotides covered at least 30×) was achieved. This allowed detection of all heterozygous germline variants in all targeted exons. Comprehensive detection of somatic mutations with a mutated gene copy frequency of 1% to 2% (wherever coverage exceeded at least 200× =99.8% of the targeted sequence) was assumed. Read depth was defined as the total number of reads at that position. Known polymorphisms identified from publically available data (http://www.ensembl.org/Homo_sapiens/) were not excluded from the analysis. Since no other biomaterials (e.g., buccal swabs) were available for the 95 MCAS patients, the germline or somatic nature of the mutations was deduced from the ratio of mutated to non-mutated alleles.

The publically available DNA NA12878 was included as a 96th sample, since a gold standard data set is available. In the gene panel counterpart, all 52 variants listed in the NIST data set for the gene panel regions were detected with no false negatives. Thus sensitivity and specificity both reached 100%. Decreasing the thresholds for variant allele frequency from default settings (25%–100%) to 1%–100% did not increase the number of called variants in this sample. This indicated that low frequency mutants detected in patients derive from somatic mutations rather than from pipeline artifacts.

Prediction of the consequences of non-synonymous single nucleotide polymorphisms (SNPs) on protein function

Two approaches were used to predict whether a given mutation may be deleterious. First, the following two algorithms tools were used to predict the impact of the detected non-synonymous SNPs on protein function: (i) sorting intolerant from tolerant (SIFT; (http://blocks.fhcrc.org/sift/SIFT.html); and (ii) polymorphisms phenotyping (PolyPhen; http://genetics.bwh.harvard.edu/pph/). The SIFT algorithm uses a query sequence to search for similar sequences that may have similar function, generates the alignment of the chosen sequences, and predicts the probability score for the impact of an amino acid substitution on protein function. SIFT scores range from 0 to 1. Outcome scores are classified as follows: 0.00–0.05, intolerant; 0.051–0.10, potentially intolerant; 0.101–0.20, borderline; and 0.201–1.00, tolerant (Ng & Henikoff 2001; Xi et al. 2004). SIFT uses information such as dbSNP ID or the GI number of an amino acid substitution to predict effects on protein function. The SIFT algorithm and

instructions for the analysis of amino acid substitutions are available at http://www.blocks.fhcrc.org/sift/SIFT.html. PolyPhen uses empirically derived rules based on previous research into protein structure, interaction, and evolution, which automatically predict whether a substitution of an amino acid is likely to be deleterious for the protein on the basis of three-dimensional structure and multiple alignments of homologous sequences (Ramensky et al. 2002; Sunyaev et al. 2001). PolyPhen input is a protein amino acid sequence or accession number, together with sequence position and two amino acid variants characterizing the polymorphism. PolyPhen outcome scores are classified as follows: 0.00–0.99, benign; 1.00–1.24, borderline; 1.25–1.49, potentially damaging; 1.50–1.99, possibly damaging; and ≥2, damaging (Xi et al. 2004). Detailed information on the Polyphen algorithm and instructions for analysis of amino acid substitutions are available at http://www.bork.embl-heidelberg.de/PolyPhen/. Second, a literature search was performed for information concerning the functional impact of the identified SNPs in cells, in particular mast cells.

Statistical analyses

Statistical analysis was complicated by two factors. First, in the reference group populations, post-hoc identification and exclusion of MCAD patients was not possible. Thus potentially, around 17% of the subjects in these groups may have been affected by MCAD (Molderings et al. 2013). This could lead to an underestimation of the differences in the allele frequencies between the study group and the reference groups. Second, if rare variants confer disease risk, an enrichment of rare causal variants is present in case groups compared to control groups if only cases are sequenced and the frequency of the detected variants is followed up in controls (Li and Leal 2009). For rare variants, this bias may be substantial, and thus the respective findings must be considered explorative. For frequent variants, no such bias should exist, thus permitting more reliable conclusions to be drawn. To compare the allele frequencies of variants between the patient group and the reference groups, Fisher's exact test was used. Here, a significance level of $\alpha = 0.05$ was set. In addition, gene based tests for the 25 candidate genes were conducted in order to test for gene-based differences in the allelic burden for cases and controls. For this purpose, the genotypes of the 1000 Genomes control sample were accessed, and the genetic regions sequenced in the present case sample were extracted. The allele counts of cases and controls within the single genes were then collapsed. Under the assumption that the collapsed mutation counts for cases and controls within genes are approximately Poisson distributed, testing was performed of the null hypothesis that the difference between the gene-based mutation rates of cases and controls (λ) equals the hypothesized difference (λ_0). In line with the single variant tests, the adjusted significance level for the gene-based tests was calculated as 0.05/25 = 0.002.

Results

Frequencies and distribution of mutational aberrations

DNA from the peripheral blood leukocytes of patients with MCAS was screened to detect non-synonymous mutations in the coding regions of 25 candidate genes. Non-synonymous mutations were detected in all 95 MCAS patients. Multiple mutations (missense, nonsense, and frameshift mutations) were detected in 19 of the 25 candidate genes (Table 3; Online Resource 2). Forty-two of the 67 non-synonymous mutations were classified by SIFT/PolyPhen as potentially deleterious/damaging. The five most frequently affected genes were *TET2*, *IL13*, *SETBP1*, *RASGRP4,* and *ASXL1*. Less frequently affected genes were *NLRP3*, *IDH1*, *IDH2*, *DNMT3A*, *KIT*, *EZH2*, *JAK2*, *CBL*, and *KMT2A*. The genes *SF3B1*, *MS4A2*, *ETV6*, *TP53,* and *RUNX1* were mutated in one or two patients only. No mutations were detected in *NRAS*, *KRAS*, *IL4*, *TNF*, *SRSF2*, or *U2AF1*. The ratio of mutated to non-mutated sequences was either close to 1 or 0.5. Therefore all of the identified mutations were considered homozygous or heterozygous germline mutations.

Besides the coding sequences (CDS), the panel covered mutational aberrations in the intronic sequences of the 25 candidate genes. The latter exceeded the number of CDS mutations many times over. However, the analysis focused on mutations in the CDS, since few data concerning the functional consequences of the intronic genetic aberrations are available.

Comparison of the frequencies of mutational aberrations in the MCAS patient group and the three reference groups

The allele frequencies of the genetic mutations detected in our in-house reference group were similar to those reported in the ExAC data and to those calculated in the sample of the 1000-Genomes Project (Table 3). Comparison of the frequency of mutations detected in the present MCAS patients with the three reference groups revealed obvious differences (Table 3). Of the 67 non-synonymous mutations found in one or more MCAS patients, 20 have not been reported in the reference groups (Table 3). The present study identified the following novel mutations: $DNMT3A^{F848L}$; $TET2^{Q745Afs*9}$; $EZH2^{R63Q}$; CBL^{D712N}; $KMT2A^{P36L}$; $KMT2A^{I3663V}$; $SETBP1^{C215*}$; $SETBP1^{G1180A}$; $SETBP1^{P1534\text{-}P1536del}$; $RASGRP4^{S519Lfs*16}$; and $RASGRP4^{C571R}$ (Table 3). The frequencies of the following mutations showed a nominally significant increase compared to the reference groups: $SF3B1^{Y141C}$; $TET2^{P363L}$; $TET2^{L1721W}$; $TET2^{I1762V}$; $TET2^{H1778R}$; $IL13^{Q144R}$; $EZH2^{R63Q}$; $KMT2A^{E533K}$; $SETBP1^{V1377L}$; $IDH2^{T352P}$;

Table 3 Non-synonymous mutations in the 25 candidate genes and their frequencies in comparison to mutational findings in the 1000 Genomes Project sample, the in-house reference group, and the ExAC data. To correct for a potential familial bias in the present study, only one randomly selected patient per family is considered in the table. Bold data indicate the presence of a significantly different allele frequency in the present cohort compared to the reference data (*p*-values of Fisher's exact test for two-sided analysis [nominal significance] shown in parantheses)

Chromosome #	Gene	rs-Number	Mutation	Genotype		Present allele frequency (n = 162)	1000 Genomes Project allele frequency (n = 1006)	inhouse allele frequency (n = 3314)	ExAC allele frequency (n = 121,412)
				homozygous	heterozygous				
1	NRAS								
1	NLRP3	rs35829419	Q705K		6	3.7%	5.1%	5.1%	4.1%
		rs147946775	S728G		1	0.6%	0.0%	0.1%	0.05%
2	IDH1	rs34218846	V178I		5	3.1%	4.8%		5.0%
		rs34599179	Y183C	1	2	2.5%	1.4%	1.4%	1.1%
2	DNMT3A	rs375399431	R676W		1	0.6%	0.0%	0.0%	0.0%
		rs147001633	R882H		1	0.6%	0.0%	0.0%	0.1%
			F848 L		1	0.6%	0.0%	0.0%	0.0%
2	SF3B1	rs199913219	Y141C		2	1.2%	**0% (p = 0.0195)**	**0.1% (p = 0.0071)**	**0.1% (p = 0.0010)**
4	KIT	rs3822214	M541 L	2	9	8.0%	7.9%		7.9%
4	TET2	rs12498609	P29R	1	4	3.7%	2.2%	3.9%	6.1%
		rs111948941	L34F		2	1.2%	1.5%	1.6%	1.5%
		rs146031219	P174H		1	0.6%	0.3%	0.4%	0.2%
		rs6843141	V218 M		3	1.9%	2.4%		5.5%
		rs377035231	H248Q		1	0.6%	0.0%	0.0%	0.01%
			Q317H		1	0.6%	0.0%	0.0%	0.0%
		rs61744960	G355D		7	4.3%	5.6%	4.6%	2.7%
		rs17253672	P363L	1	14	9.9%	7.0%	6.0%	**4.8% (p = 0.0046)**
		rs201642693	G429R		1	0.6%	0.0%	0.0%	0.04%
			M533I		1	0.6%	0.0%	0.0%	0.0%
			Q745Afs*9		1	0.6%	0.0%	0.0%	0.0%
		rs144386291	Y867H		2	1.2%	0.5%	0.7%	0.7%
		rs75056899	Q1084P		1	0.6%	0.7%	0.2%	0.3%
		rs34402524	L1721 W	3	17	14.2%	14.0%		12.9%
		rs146348065	P1723S		2	1.2%	0.5%	0.7%	0.6%
		rs2454206	I1762V	13	37	38.9%	37.2%		**29.5% (p = 0.0113)**
		rs62621450	H1778R		3	1.9%	**0% (p = 0.0027)**		4.9%
5	IL13	rs20541	Q144R	57	24	85.2%	79.3%		**73.3% (p = 0.0005)**
7	EZH2		R63Q	1	1	1.9%	**0% (p = 0.0027)**	**0.1% (p = 0.0001)**	0.0%
		rs2302427	D185H		18	11.1%	7.0%		7.9%
9	JAK2	rs142269166	N1108S		2	1.2%	0.2%	0.2%	**0.2% (p = 0.0397)**

Table 3 (continued)

Chromosome #	Gene	rs-Number	Mutation	Genotype		Present allele frequency (n = 162)	1000 Genomes Project allele frequency (n = 1006)	inhouse allele frequency (n = 3314)	ExAC allele frequency (n = 121,412)
				homozygous	heterozygous				
		rs41316003	R1063H		2	1.2%	0.4%	0.6%	0.4%
11	CBL	rs373212940	H42_L43insH		1	0.6%	0.1%	0.0%	0.0%
		rs727504504	Q367K		1	0.6%	0.0%	0.0%	0.0%
		rs369030902	D549E		1	0.6%	0.0%	0.0%	0.0%
			D712N		1	0.6%	0.0%	0.0%	0.0%
11	MS4A2	rs569108	E237G		4	2.5%	3.7%		4.8%
11	KMT2A	rs9332745	A30G		1	0.6%	0.0%	0.1%	1.3%
			P36L		1	0.6%	0.0%	0.0%	0.0%
			E535K		3	1.9%	**0% (p = 0.0027)**	0.0%	0.0%
		rs138969270	L3134F		1	0.6%	0.0%	0.1%	0.0%
			I3663V		1	0.6%	0.0%	0.0%	0.0%
12	KRAS								
12	ETV6	rs145477191	L201P		1	0.6%	1.2%	0.4%	0.5%
15	IDH2	rs118101777	R261H		1	0.6%	0.5%	0.3%	0.2%
		rs770548230	T352P	1		1.2%	**0% (p = 0.0195)**		
		rs118053940	T435 M		1	0.6%	0.6%	0.7%	0.4%
17	SRSF2								
17	TP53	rs1042522	P72R	52	24	79.0%	71.5%		**66% (p = 0.0012)**
18	SETBP1		C215*		1	0.6%	0.0%	0.0%	0.0%
		rs663651	A222T	23	39	52.5%	56.2%		50.7%
		rs3085861	T228Sfs*8	22	39	51.2%	56.2%		coupled with A222T
		rs11082414	V231 L	6	28	24.7%	20.3%		19.3%
		rs3744825	V1101I	2	18	13.6%	13.2%		12.1%
		rs1064204	P1130T	1	13	9.3%	11.0%		12.9%
			G1180A		1	0.6%	0.0%	0.0%	0.0%
		rs77518617	V1377 L		4	2.5%	1.7%	1.1%	**0.7% (p = 0.0262)**
			P1534_P1536del		1	0.6%	0.0%	0.0%	0.0%
19	RASGRP4	rs892055	I18T	10	34	33.3%	35.5%		43.1%
		rs41465445	G165R		6	3.7%	4.2%	5.1%	4.2%
		rs202008979	R335G	1		1.2%	0.6%	0.8%	0.5%
			S519Lfs*16	1		1.2%	**0% (p = 0.0195)**	0.0%	0.0%
			C571R		1	0.6%	0.0%	0.0%	0.0%
		rs34377632	E620K		4	2.5%	3.4%	3.4%	4.0%

Table 3 (continued)

Chromosome #	Gene	rs-Number	Mutation	Genotype		Present allele frequency (n = 162)	1000 Genomes Project allele frequency (n = 1006)	inhouse allele frequency (n = 3314)	ExAC allele frequency (n = 121,412)
				homozygous	heterozygous				
		rs117355762	R638H		1	0.6%	1.0%	0.6%	1.7%
20	ASXL1	rs3746609	G652S		1	0.6%	0.5%	1.9%	0.0%
		rs145132837	N986S	1		1.2%	0.2%	0.3%	0.1%
		rs139115934	E1102D		4	2.5%	2.1%	1.3%	1.0%
		rs146464648	G1397S		1	0.6%	0.3%	0.3%	0.2%
21	U2AF1								
21	RUNX1	rs111527738	L56S		2	1.2%	1.7%	1.4%	1.6%

JAK2^{N1108S}; RASGRP4^{S519Lfs*16}; and TP53^{P72R} (Table 3, bold data).

Comparison of the overall burden of mutations in the present MCAS cohort with those reported in the 1000 Genomes Project sample revealed no significant difference. The result remained negative when the comparison was made at the level of single genes (Online Resource 3).

A comparison of gene-wise average coverage was performed. This excluded the presence in our cohort of larger copy number variations involving any of the 25 candidate genes.

Comparison with KIT mutations identified previously in 13 of the present MCAS patients

Table 4 lists KIT mutational alterations detected in 13 of the present MCAS patients in a previous study by our group using mRNA from isolated mast cell progenitor cells (Molderings et al. 2010). With the exception of KIT^{M541L} in patient #48 (Table 4), none of these genetic alterations were detected in the present NGS study using DNA from peripheral blood leukocytes. The previous study also identified a single JAK2 variant (JAK2^{R1063H}) in one patient. In the present study, this variant was detected in the same patient as a germline mutation.

Familial co-segregation of the detected variants

Of the 42 rare variants (allele frequency < 1% in the 1000 Genomes Project) detected in the present MCAS cohort, four were observed at least once in our multiply affected families. Three of these variants co-segregated with disease in smaller families. However, this may have been attributable to chance. In our single large pedigree, only one rare variant was detected, and this did not show co-segregation. This large pedigree included six affected individuals across two generations (Fig. 1).

Discussion

Somatic mutations in the selected genes in MCAS patients

To date, most genetic studies of MCAD have focused on SM (for review, see Molderings 2015). Besides the presence of the mutation KIT^{D816V}, somatic mutations have been identified in a total of 25 other genes in leukocytes from the peripheral blood of SM patients (Schwaab et al. 2013; Traina et al. 2012; Damaj et al. 2014; Rechsteiner et al. 2014; Jawhar et al. 2015, 2016). The first aim of the present study was to investigate whether patients with MCAS also carry mutations in any of these 25 candidate genes in DNA isolated from peripheral blood leucocytes. This was achieved using NGS.

Table 4 Results of a previous molecular genetic analysis of *KIT* transcripts by the present authors (Molderings et al. 2010). Mutations also detected in the present study are shown in bold

Patient #	Alterations in KIT mRNA
27	Ins Q253 Del 510–513 Del S715 D760V Del nt C2311(codon 764) ⇒ frame shift with stop codon at amino acid 764 Ins of 63 nucleotides from the Est AF95853 at nt2104 Ins of 48 nucleotides from the Est AF95853 at nt2673
28	Ins Q253 Del 510–513 Del S715 I797S R804W D816G
40	E53K Ins Q252 Del GNNK510–513 Del S715 Complex alteration between amino acids 343–519 nt g243a (Codon 81) silent mutation Insertions at the C-terminus
42	Del 510–513 Stop codon at amino acid 475 Del S715
43	Ins Q253 ins380 unknown sequence Del 510–513 K558R Del S715 E720K
48	Ins Q253 Del 510–513 M541 L Del S715
50	Ins Q253 Del 510–513 Del nt A2356 (codon 778) ⇒ frame shift with stop codon at amino acid 814
52	Ins Q253 Del 510–513 Del S715 nt c70t silent mutation
62	Ins Q252 E270K D327N E338K M351I Del 510–513 Del S715 nt t2094c (codon 691) silent mutation (homozygous)
67	Del 510–513 Del S715
71	E53K K116 N Ins Q253 G285R Del 510–513 Del S715 nt c2509t ⇒ 830 stop codon nt c621t (codon 200) silent mutation nt c645t (codon 208) silent mutation nt t2205c (codon 728) silent mutation
73	Ins Q253 Del 510–513 M541 L Y672S E554K Del S715 nt c2608g (codon 870) silent mutation
83	E53K E73R T74R Ins Q252 (+/− 0.9%) Del GNNK510–513 M541 L Del S715 (−/+ 50.8%) nt g2607c (Codon 869) silent mutation Weak insertions at the C-terminus

In the present 95 MCAS patients, none of the somatic mutations reported by Schwaab et al. 2013 and Hanssens et al. 2014 as being of potential etiological relevance to SM were detected, and neither were any other somatic mutations in the respective genes. This discrepancy has several potential explanations. First, the present analyses were conducted in DNA from peripheral blood leukocytes. In contrast, Hanssens et al. (2014) investigated leukocytes from bone marrow, while Schwaab et al. 2013 used samples from both peripheral blood (n = 32) and bone marrow (n = 9). Second, patients in the Schwaab et al. 2013 and Hanssens et al. 2014 studies presented with advanced forms of SM. Therefore, their tissue samples may have harbored a considerable number of mast cells and mast cell progenitor cells. Furthermore, multilineage mutation spread has been reported for advanced forms of SM. Hence, the amount of mast cell DNA present in total leukocyte DNA is likely to have been much higher than was the case in our samples. This dilution is likely to have shifted the identification of somatic mutations below the detection limit of the NGS technique. An alternative explanation is that some of these mutations emerged during the transcription process, as a result of germline mutations in regulatory genes. In a previous study by our group, mutations in the cDNA, i.e., at the RNA level, were detected in the mast cells of MCAS patients (Molderings et al. 2007, 2010). However, no investigations were performed using mast cell genomic DNA. Furthermore, many of the mutations were alterations at splicing sites, and were thus likely to have been attributable to alternative splicing and/or splicing errors that could not be detected in the present NGS step at the level of genomic DNA.

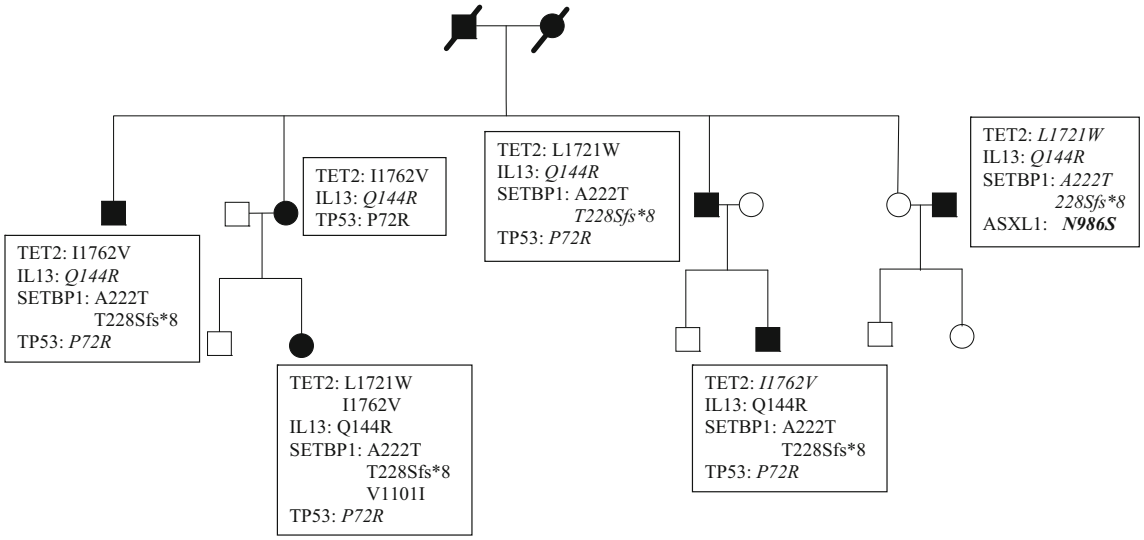

Fig. 1 Pedigree with a high familial loading of mast cell activation syndrome. Squares, males; circles, females. Filled symbol, patient affected with MCAS; open symbol, clinically healthy individual. Slash, deceased. Bold type, rare variant. Italics, homozygous mutations

Taken together, these findings suggest that somatic mutations that are present exclusively in mast cells cannot be detected via NGS of DNA from peripheral blood leukocytes, except when the number of circulating mast cells and mast cell-committed progenitor cells is significantly increased and/or multilineage mutation spread has already occurred.

Germline mutations in the 25 candidate genes in MCAS patients

Although authors of previous NGS studies of SM have assumed that the detected mutations were somatic in nature, the criteria applied in this assessment were unclear (Schwaab et al. 2013; Traina et al. 2012; Damaj et al. 2014, Hanssens et al. 2014). In the present study, the ratio of sequence reads of all detected mutations was close to 1 for homozygous mutations and close to 0.5 for heterozygous mutations. In view of this, and the fact that the DNA was isolated from a heterogeneous cell population of hematopoietic origin in peripheral blood, we conclude that the detected mutations were germline in nature. Their germline character is further supported by the findings in the seven different pedigrees, one of which included a monozygotic twin pair (Fig. 1, Online Resource 4). In these pedigrees, all mutations were present in at least one of the two parents, when both parents were available for investigation, thus rendering a somatic nature unlikely.

Potential relevance of the detected germline mutations to mast cell dysfunction in MCAS patients

In the allelic burden analysis, no significant gene-based differences were found, indicating that there is no large effect, at least within the 25 candidate genes of interest in the present analyses. However, this finding does not exclude the possibility of smaller effects for single genes. If such effects are present, this may mean not only that the effect sizes are too small to result in a significant gene-based analysis but also, for example, that only a small fraction of the mutations occurring in a particular gene are associated with disease development. In fact, for several of the investigated genes, individual mutations showed nominally significant differences between patients and controls. As mentioned above, these results must be interpreted with caution, since the present study design (sequencing of patients and follow-up of detected variants in controls) produces a bias towards false positive results. With this limitation in mind, the association findings provide support for some of the investigated genes, which may warrant follow up in larger samples.

Among the frequent variants, the largest observed differences between patients and controls were found for $IL13^{Q144R}$ and $TP53^{P72R}$. IL13 encodes interleukin 13, an immunoregulatory cytokine which is generated primarily by activated Th2 cells. The variant Q144R was detected in 94 of the 95 MCAS patients, with an allele frequency of 85.2% compared to 79.3% in controls. Genome-wide association studies have repeatedly demonstrated that this polymorphism, i.e., the R-allele, is associated with both IgE dysregulation (Granada et al. 2012; references listed at http://snpedia.com/index.php/Rs20541), and classical Hodgkin's lymphoma (Urayama et al. 2012; references listed at http://snpedia.com/index.php/Rs20541). A plausible hypothesis is that the Q144R variant predisposes to MCAD by enhancing functional activation via IgE dysregulation and/or other unknown pathological processes.

TP53 encodes the major tumor suppressor protein p53. The p53 protein contains transcriptional activation-, DNA binding-, and oligomerization domains, which are involved in various biological processes, including the control of cell-cycle

checkpoints and apoptosis. Both SIFT and Polyphen classify the sequence variant P72R as deleterious and possibly damaging. The P72R polymorphism, i.e., the R-allele, has been reported to be associated with an increased risk for acute lymphoblastic leukemia (Do et al. 2009); endometrial cancer (Ashton et al. 2009); breast cancer (Johnson et al. 2007; Rajkumar et al. 2008); and mycosis fungoides (McGirt et al. 2015).

A limitation of the present study was that controls from the in-house reference group and the 1000 Genomes Project sample were not screened for MCAD. This may have impacted our results, since the prevalence of MCAD in the German general population is around 17%, and this is mainly attributable to MCAS (Molderings et al. 2013). Furthermore, the ExAc data were collected from patients with diseases that are associated with MCAS (Online Resource 5). Hence, an enrichment of variants with a causal relationship to MCAS might have been present in the ExAc reference group. The use of unscreened controls in the in-house reference group and the 1000 Genomes Project sample, and a possible enrichment of patients with MCAS-associated diseases in the ExAc reference group, may have biased the statistical analysis towards less significant differences.

Conclusions

The present analyses detected no somatic mutations in genes implicated in recent studies of SM, despite high NGS coverage. This failure is probably attributable to the dilution of mutated mast cell DNA in normal leukocyte DNA. However, multiple non-synonymous germline mutations were detected in the coding regions of the 25 candidate genes. Overlap with mutations identified previously in SM was small. Some of these variants may be functionally relevant (Online Resource 6) and the present exploratory association findings suggest that these are interesting candidate variants for follow up studies.

Acknowledgements We thank Dr. Christine Schmäl for her linguistic revision of the manuscript. This study was supported by a grant from the *Förderclub Mastzellforschung e.V., Germany.*

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the ethics committee of the Medical Faculty of the University of Bonn (approval no. 144/14).

Informed consent All participants received detailed information concerning the purpose of the study, and provided written informed consent prior to inclusion. Informed consent was obtained from all individual participants included in the study.

References

1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526(7571):68–74

Afrin LB, Butterfield J, Raithel M, Molderings GJ (2016) Mast cell activation disease: often seen, rarely recognized. Ann Med 48:190–201

Akin C, Valent P, Metcalfe D (2010) Mast cell activation syndrome: proposed diagnostic criteria. J Allergy Clin Immunol 126:1099–1104

Alfter K, von Kügelgen I, Haenisch B, Frieling T, Hülsdonk A, Haars U, Rolfs A, Noe G, Kolck UW, Homann J, Molderings GJ (2009) New aspects of liver abnormalities as part of the systemic mast cell activation syndrome. Liver Int 29:181–186

Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, Attia J, Gilbert M, Hamann U, Scott RJ (2009) Polymorphisms in TP53 and MDM2 combined are associated with high grade endometrial cancer. Gynecol Oncol 113:109–114

Damaj G, Joris M, Chandesris O, Hanssens K, Soucie E, Canioni D, Kolb B, Durieu I, Gyan E, Livideanu C, Chèze S, Diouf M, Garidi R, Georgin-Lavialle S, Asnafi V, Lhermitte L, Lavigne C, Launay D, Arock M, Lortholary O, Dubreuil P, Hermine O (2014) ASXL1 but not TET2 mutations adversely impact overall survival of patients suffering systemic mastocytosis with associated clonal hematologic non-mast-cell diseases. PLoS One 9:e85362

Do TN, Ucisik-Akkaya E, Davis CF, Morrison BA, Dorak MT (2009) TP53 R72P and MDM2 SNP309 polymorphisms in modification of childhood acute lymphoblastic leukemia susceptibility. Cancer Genet Cytogenet 195:31–36

Garcia-Montero AC, Jara-Acevedo M, Teodosio C, Sanchez ML, Nunez R, Prados A, Aldanondo I, Sanchez L, Dominguez M, Botana LM, Sanchez-Jimenez F, Sotlar K, Almeida J, Escribano L, Orfao A (2006) KIT mutation in mast cells and other bone marrow hematopoietic cell lineages in systemic mast cell disorders: a prospective study of the Spanish network on Mastocytosis (REMA) in a series of 113 patients. Blood 108:2366–2372

Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, Albrecht E, Gieger C, Heinrich J, Himes BE, Hunninghake GM, Celedón JC, Weiss ST, Cruikshank WW, Farrer LA, Center DM, O'Connor GT (2012) A genome-wide association study of plasma total IgE concentrations in the Framingham heart study. J Allergy Clin Immunol 129:840–845

Hamilton MJ, Hornick JL, Akin C, Castells MC, Greenberger NJ (2011) Mast cell activation syndrome: a newly recognized disorder with systemic clinical manifestations. J Allergy Clin Immunol 128:147–152

Hanssens K, Brenet F, Agopian J, Georgin-Lavialle S, Damaj G, Cabaret L, Chandesris MO, de Sepulveda P, Hermine O, Dubreuil P, Soucie E (2014) SRSF2-p95 hotspot mutation is highly associated with advanced forms of mastocytosis and mutations in epigenetic regulator genes. Haematologica 99:830–835

Hermine O, Lortholary O, Leventhal PS, Catteau A, Soppelsa F, Baude C, Cohen-Akenine A, Palmérini F, Hanssens K, Yang Y, Sobol H, Fraytag S, Ghez D, Suarez F, Barete S, Casassus P, Sans B, Arock M, Kinet JP, Dubreuil P, Moussy A (2008) Case-control cohort study of patients' perceptions of disability in mastocytosis. PLoS One 3:e2266

Jawhar M, Schwaab J, Schnittger S, Sotlar K, Horny HP, Metzgeroth G, Müller N, Schneider S, Naumann N, Walz C, Haferlach T, Valent P, Hofmann WK, Cross NC, Fabarius A, Reiter A (2015) Molecular profiling of myeloid progenitor cells in multi-mutated advanced systemic mastocytosis identifies KIT D816V as a distinct and late event. Leukemia 29:1115–1122

Jawhar M, Schwaab J, Schnittger S, Meggendorfer M, Pfirrmann M, Sotlar K, Horny HP, Metzgeroth G, Kluger S, Naumann N, Haferlach C, Haferlach T, Valent P, Hofmann WK, Fabarius A,

Cross NC, Reiter A (2016) Additional mutations in SRSF2, ASXL1 and/or RUNX1 identify a high-risk group of patients with KIT D816V(+) advanced systemic mastocytosis. Leukemia 30:136–143

Johnson N, Fletcher O, Palles C, Rudd M, Webb E, Sellick G, dos Santos SI, McCormack V, Gibson L, Fraser A, Leonard A, Gilham C, Tavtigian SV, Ashworth A, Houlston R, Peto J (2007) Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. Hum Mol Genet 16:1051–1057

Kristensen T, Broesby-Olsen S, Vestergaard H, Bindslev-Jensen C, Møller MB, on behalf of the Mastocytosis Centre Odense University Hospital (MastOUH) (2012) Circulating KIT D816V mutation positive non-mast cells in peripheral blood are characteristic of indolent systemic mastocytosis. Eur J Haematol 89:42–46

Li B, Leal SM (2009) Discovery of rare variants via sequencing: implications for the design of complex trait association studies. PLoS Genet 5:e1000481

Longley BJ, Metcalfe DD, Tharp M, Wang X, Tyrrell L, Lu SZ, Heitjan D, Ma Y (1999) Activating and dominant inactivating c-kit catalytic domain mutations in distinct clinical forms of human mastocytosis. Proc Natl Acad Sci 96:1609–1614

McGirt LY, Jia P, Baerenwald DA, Duszynski RJ, Dahlman KB, Zic JA, Zwerner JP, Hucks D, Dave U, Zhao Z, Eischen CM (2015) Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. Blood 126:508–519

Molderings GJ (2015) The genetic basis of mast cell activation disease - looking through a glass darkly. Crit Rev Oncol Hematol 93:75–89

Molderings GJ, Kolck UW, Scheurlen C, Brüss M, Homann J, Von Kügelgen I (2007) Multiple novel alterations in kit tyrosine kinase in patients with gastrointestinally pronounced systemic mast cell activation disorder. Scand J Gastroenterol 42:1045–1053

Molderings GJ, Meis K, Kolck UW, Homann J, Frieling T (2010) Comparative analysis of mutation of tyrosine kinase kit in mast cells from patients with systemic mast cell activation syndrome and healthy subjects. Immunogenetics 62:721–727

Molderings GJ, Brettner S, Homann J, Afrin LB (2011) Mast cell activation disease: a concise practical guide for diagnostic workup and therapeutic options. J Hematol Oncol 4:10

Molderings GJ, Haenisch B, Bogdanow M, Fimmers R, Nöthen MM (2013) Familial occurrence of systemic mast cell activation disease PLoS One 8:e76241

Nagata H, Worobec AS, Oh CK, Chowdhury BA, Tannenbaum S, Suzuki Y, Metcalfe DD (1995) Identification of a point mutation in the catalytic domain of the protooncogene c- kit in peripheral blood mononuclear cells of patients who have mastocytosis with an associated hematologic disorder. Proc Natl Acad Sci 92:10560–10564

Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11:863–874

Rajkumar T, Samson M, Rama R, Sridevi V, Mahji U, Swaminathan R, Nancy NK (2008) TGFbeta1 (Leu10Pro), p53 (Arg72Pro) can predict for increased risk for breast cancer in south Indian women and TGFbeta1 pro (Leu10Pro) allele predicts response to neo-adjuvant chemo-radiotherapy. Breast Cancer Res Treat 112:81–87

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900

Rechsteiner M, Müller R, Reineke T, Goede J, Bohnert A, Zhong Q, Manz MG, Moch H, Wild PJ, Zimmermann DR, Tinguely M (2014) Modelling of a genetically diverse evolution of systemic Mastocytosis with chronic Myelomonocytic leukemia (SM-CMML) by next generation sequencing. Exp Hematol Oncol 3:18

Schwaab J, Schnittger S, Sotlar K, Walz C, Fabarius A, Pfirrmann M, Kohlmann A, Grossmann V, Meggendorfer M, Horny HP, Valent P, Jawhar M, Teichmann M, Metzgeroth G, Erben P, Ernst T, Hochhaus A, Haferlach T, Hofmann WK, Cross NC, Reiter A (2013) Comprehensive mutational profiling in advanced systemic mastocytosis. Blood 122:2460–2466

Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P (2001) Prediction of deleterious human alleles. Hum Mol Genet 10: 591–597

Teodosio C, García-Montero AC, Jara-Acevedo M, Alvarez-Twose I, Sánchez-Muñoz L, Almeida J, Morgado JM, Matito A, Escribano L, Orfao A (2012) An immature immunophenotype of bone marrow mast cells predicts for multilineage D816V KIT mutation in systemic mastocytosis. Leukemia 26:951–958

Traina F, Visconte V, Jankowska AM, Makishima H, O'Keefe CL, Elson P, Han Y, Hsieh FH, Sekeres MA, Mali RS, Kalaycio M, Lichtin AE, Advani AS, Duong HK, Copelan E, Kapur R, Olalla Saad ST, Maciejewski JP, Tiu RV (2012) Single nucleotide polymorphism array lesions, TET2, DNMT3A, ASXL1 and CBL mutations are present in systemic mastocytosis. PLoS One 7:e43090

Urayama KY, Jarrett RF, Hjalgrim H, Diepstra A, Kamatani Y, Chabrier A, Gaborieau V, Boland A, Nieters A, Becker N, Foretova L, Benavente Y, Maynadié M, Staines A, Shield L, Lake A, Montgomery D, Taylor M, Smedby KE, Amini RM, Adami HO, Glimelius B, Feenstra B, Nolte IM, Visser L, van Imhoff GW, Lightfoot T, Cocco P, Kiemeney L, Vermeulen SH, Holcatova I, Vatten L, Macfarlane GJ, Thomson P, Conway DI, Benhamou S, Agudo A, Healy CM, Overvad K, Tjønneland A, Melin B, Canzian F, Khaw KT, Travis RC, Peeters PH, González CA, Quirós JR, Sánchez MJ, Huerta JM, Ardanaz E, Dorronsoro M, Clavel-Chapelon F, Bueno-de-Mesquita HB, Riboli E, Roman E, Boffetta P, de Sanjosé S, Zelenika D, Melbye M, van den Berg A, Lathrop M, Brennan P, McKay JD (2012) Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. J Natl Cancer Inst 104:240–253

Valent P, Akin C, Escribano L, Fodinger M, Hartmann K, Brockow K, Castells M, Sperr WR, Kluin-Nelemans HC, Hamdy NA, Lortholary O, Robyn J, van Doormaal J, Sotlar K, Hauswirth AW, Arock M, Hermine O, Hellmann A, Triggiani M, Niedoszytko M, Schwartz LB, Orfao A, Horny HP, Metcalfe DD (2007) Standards and standardization in mastocytosis: consensus statements on diagnostics, treatment recommendations and response criteria. Eur J Clin Investig 37:435–453

Valent P, Akin C, Arock M, Brockow K, Butterfield JH, Carter MC, Castells M, Escribano L, Hartmann K, Lieberman P, Nedoszytko B, Orfao A, Schwartz LB, Sotlar K, Sperr WR, Triggiani M, Valenta R, Horny HP, Metcalfe DD (2012) Definitions, criteria and global classification of mast cell disorders with special reference to mast cell activation syndromes: a consensus proposal. Int Arch Allergy Immunol 157:215–225

Xi T, Jones IM, Mohrenweiser HW (2004) Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics 83: 970–979