Biochimica et Biophysica Acta 1859 (2016) 1411–1428

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Mechanism of allele specific assembly and disruption of master regulator transcription factor complexes of NF-κBp50, NF-κBp65 and HIF1a on a non-coding FAS SNP

Chidiebere U. Awah [a,b], Stephanie Tamm [a], Silke Hedtfeld [a], Doris Steinemann [c], Burkhard Tümmler [a,d], Georgios Tsiavaliaris [e], Frauke Stanke [a,d,*]

[a] Department of Paediatric Pneumology, Neonatology and Allergology, Hannover Medical School, Hannover, Germany
[b] Graduate School of Excellence, MD/PhD Programme Molecular Medicine Hannover Biomedical Research School, Hannover Biomedical Research School, Hannover Medical School, Hannover, Germany
[c] Institute for Human Genetics, Hannover Medical School, Hannover, Germany
[d] Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Centre for Lung Research, Germany
[e] Institute for Biophysical Chemistry, Hannover Medical School, Hannover, Germany

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form 22 August 2016
Accepted 7 September 2016
Available online 9 September 2016

Keywords:
NF-κBp50
NF-κBp65
HIF1a
Higher order multimeric TF complexes
Allele-specific TF binding
TF-binding affinity
Non-coding SNP
Cystic fibrosis

ABSTRACT

A challenging question in genetics is to understand the molecular function of non-coding variants of the genome. By using differential EMSA, ChIP and functional genome analysis, we have found that changes in transcription factors (TF) apparent binding affinity and dissociation rates are responsible for allele specific assembly or disruption of master TFs: we observed that NF-κBp50, NF-κBp65 and HIF1a bind with an affinity of up to 10 fold better to the C-allele than to the T-allele of rs7901656 both *in vivo* and *in vitro*. Furthermore, we showed that NF-κBp50, p65 and HIF1a form higher order heteromultimeric complexes overlapping rs7901656, implying synergism of action among TFs governing cellular response to infection and hypoxia. With rs7901656 on the *FAS* gene as a paradigm, we show how allele specific transcription factor complex assembly and disruption by a causal variant contributes to disease and phenotypic diversity. This finding provides the highly needed mechanistic insight into how the molecular etiology of regulatory SNPs can be understood in functional terms.

1. Introduction

Non-coding single nucleotide polymorphisms (SNPs) account for more than 90% of all known SNPs identified by genome wide association studies (GWAS), which has linked them to complex human traits and disease susceptibility loci [1–3]. Disruption and variation in transcription factor binding [4–14] and changes in chromatin states [15–16] on risk variants have been shown to predispose to colorectal cancers, alpha-thalassaemia, dyslipidemia and autoimmune diseases but the understanding of the underlying molecular mechanisms responsible for disruption of transcription factor binding by non-coding variants and subsequent involvement in disease development or manifestation has remained elusive.

Regulatory SNPs are involved in the regulation of variation of gene expression; they have been hypothesized to be the source of phenotypic variability and basis of complex diseases [17–19]. Interest on the importance of these SNPs was ignited when it was proven that the number of sequences encoding for proteins in the human genome was surprisingly low, approximately 1.5%, with non-coding elements and regulatory DNA accounting for the remaining proportion [20]. This was further supported by GWAS which has shown that most SNPs associated with diseases lie on non-coding regions [41]. Due to their location on the intergenic regions, it has been rather difficult to establish their functions [18-21]. Initial *in silico* analysis of these non-coding SNPs indicated that there is high a probability that transcriptions factors such as MEF2, (a master regulator), bind to these regions [22]. Although this initial finding was entirely based on *in silico* methods, it has given directions that transcription factors, their binding to variable SNPs and subsequent impact on gene regulation might be important for the pathophysiology

* Corresponding author at: Department of Paediatric Pneumology, Neonatology and Allergology, OE6710, Carl-Neuberg-Strasse 1, Hannover Medical School, 30625 Hannover, Germany.
E-mail address: Mekus.Frauke@mh-hannover.de (F. Stanke).

http://dx.doi.org/10.1016/j.bbagrm.2016.09.002
1874-9399/

of complex diseases. A putative mechanism of how a TF interaction with an SNP might be involved in gene regulation has been suggested by Buckland, 2006 [17]. One of such mechanisms is an interaction between a TF and an SNP which will induce changes in DNA structure. In these circumstances, transcriptional output may be favorably or unfavorably perturbed, leading to changes in gene expression [17].

Nuclear factor kappa B (NF-κB) is a master regulator transcription factor because of its central role in immune and inflammatory response regulation and its involvement in adaptive and innate immunity, embryonic development, infections, apoptosis and oncogenesis [23–30]. NF-κB binds and activates transcriptional regulation of its target genes as a heterodimer NF-κBp50:p65 [31].

Hypoxia inducible factor 1a (HIF1a) is also a master regulator transcription factor responsible for controlling oxygen sensing, erythropoiesis, metabolism and angiogenesis both in healthy and cancerous cells [32–34]. HIF1a belongs to the mammalian transcription factor family of basic-helix–loop–helix PER-ARNT-SIM (bHLH PAS) [35] which includes the aryl hydrocarbon receptor (AHR). This TF family is mainly responsible for sensing environmental cues and driving transcriptional responses to meet cellular needs [34]. HIF1a::ARNT is the transcriptional active form of HIF1a [35].

FAS is a transmembrane protein that belongs to the tumour necrosis factor receptor family. The binding of its ligand FASL leads to the induction of programmed cell death (apoptosis) in many cell types. FASL/FAS interaction orchestrates important functions in immune homeostasis, surveillance and regulation of inflammation [36]. FASL has been shown to be very important in host defense and modulation of host–pathogen interaction during *Pseudomonas aeruginosa* infection in the lung [37], by inducing apoptosis in *P. aeruginosa* and FAS infected cells and may activate NF-κB for cytokine secretion from the epithelium which kills the extracellular bacteria and protects the lung tissue [37]. An association study of twins and siblings with Cystic Fibrosis (CF) identified the FAS SNP rs7901656 in the intron 2 of the *FAS* gene as a clinically meaningful genetic modifier of CF, with two contrasting alleles C or T associating with mild and severe clinical outcome respectively [38]. However, the conventional method of sequencing showed neither an alteration in the coding sequences nor at splice sites in linkage disequilibrium to rs7901656. This indicated that an unknown gene regulatory mechanism may be causal for the observed association of lower FAS expression levels and mild CF disease observed among carriers of the rs7901656 C-allele [38].

In this study, we set out to find the mechanistic basis of how the two alleles of rs7901656 on the non-coding region of FAS regulate FAS expression and how FAS expression modifies CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) mediated residual chloride ion transport which leads to the differential cystic fibrosis clinical outcome observed in patients. To accomplish this aim we performed broad functional analysis and found that the master transcription factors NF-κB and HIF1a preferentially bind the C-allele of rs7901656. We probed how these master transcription factor complexes are assembled or disrupted on the alleles of rs7901656. We found the minimum sequence motifs required for the transcription factors to recognize their binding sites for efficient binding and studied the influence of the base substitution (C versus T) as well as the effect of the surrounding sequences on transcription factor complex formation. We determined equilibrium dissociation constants (K_{dapp}), rates and binding constants of the allele-TF association (k_{obs}, k_{on}), and dissociation rates (k_{off}) of the complexes. Our data show that NF-κBp50 and HIF1a favour C-allele binding with high affinity and form higher order multimeric complexes which indicate that C-allele specific TF interaction is responsible for the increased formation and stability of the complex. This provides an explanation of how causal genetic variants that differ in a single base pair influence cystic fibrosis disease development. We further used chromatin immunoprecipitation and found that NF-κBp50, NF-κBp65 and HIF1a preferentially bind the C-allele in vivo across four genotyped cell lines with contrasting rs7901656 genotypes. This provides evidence of synergism among TFs governing cellular response to inflammation and hypoxia. We performed an association study with four hundred patient samples comprising 101 families with siblings homozygous for F508del-CFTR and found that variation in *NF-κB1* encoding for NF-κBp50 associates with CF disease manifestation. Thus, we demonstrate with our studies on rs7901656 in *FAS* how alleles on non-coding regions promote disease and modify disease outcome.

2. Materials and methods

2.1. Functional genome in silico analysis

We used complementary analytical approaches to identify functional elements (transcription factors, transcription factor binding sites, RNA PolII, histone marks, nucleosome occupancy) on or around the SNP rs7901656 with ENCODE and MOTIFMAP data. We investigated the preferential binding of transcription factors on the C vs T-allele of rs7901656 exchange by JASPAR and we used RSAT and sequence alignment by Clustal Omega [97] to determine how NF-κB complex subunits (p50, p65) and HIF1a are assembled or disrupted on the C or T-allele of rs7901656. For detailed description of each computational approach used, see the Supplementary methods.

2.2. Extraction of nuclear proteins

The immortalized human bronchial cell line 16HBE14o- and the cancer intestinal epithelium cell line T84, both unstimulated, were grown to confluence, snap frozen in liquid Nitrogen and stored at −80 °C. Nuclear extracts were prepared according to [98] with a few modification to the method. A detailed description is provided in the Supplementary methods.

2.3. Probes

31 bp oligonucleotides surrounding the C-allele of rs7901656, the T-allele of rs7901656 and a scrambled control probe were used. The scrambled probe contains a consensus sequences of NF-κB which was changed to its complementary sequence at positions 1 (A>T), 3 (G>C), 4(A>T), 7 (T>A) and 10 (C>G. The Scr probe served as a measure to evaluate NF-κB binding on the C- and the T-allele. Moreso, a HIF1a response element was introduced by serendipity on the Scr probe, weight score 2.0, which served as a positive control for HIF1a binding (Fig. 1G). All probes were biotinylated at the 5′ end and at the 3′ end on the reverse strand, synthesized by Biotez Berlin–Buch GmbH, HPLC purified, at 40 nmol synthesis scale. Final concentrations used in all experiments were 50 μM except in the serial dilution experiments where concentrations ranging from 0.39 μM to 50 μM were used and in the rate kinetic experiment where 25 μM of the probes was used. All probes used in experiments are listed in the Supplementary methods.

2.4. Electrophoretic mobility shift (EMSA) and supershift assays

Both complementary strands of DNA probes C-allele, T-allele, NF-κB p50 consensus probe, NF-κB p65 consensus probe and Scr probe were denatured in EMSA binding buffer A at 75 °C for 5 min and slowly cooled overnight. Probes were incubated with binding buffer B at 37 °C for 1 h (for composition of buffer A and B, see Supplementary methods). Master mix (80 μM annealed poly dI–dC, glycerin and chicken liver genomic DNA (1400 ng/μl) was added and incubated at 37 °C for 1 h. The 16HBE14o- nuclear extract (20 ng of protein) was added and incubated at 37 °C for 2 h. Antibodies were added and incubated at 37 °C for 1 h. A graphical depiction of the EMSA method used is shown in S2A Fig. A detailed description of each EMSA obtained and accompanying figures and replicates are shown in the supplementary methods. All antibodies and