Identifying hidden drivers of heterogeneous inflammatory diseases

Authors: Natalie Garzorz-Stark[1,2,3,†], Richa Batra[3,,8*,†], Felix Lauffer[3,†], Manja Jargosch[3,5†], Caroline Pilz[3], Sophie Roenneberg[3], Alexander Schäbitz[1,3], Alexander Böhner[3], Peter Seiringer[3], Jenny Thomas[5], Bentolhoda Fereydouni[1], Lam C Tsoi[6], Johann E Gudjonsson[6], Fabian Theis F[4,7], Tilo Biedermann[3], Carsten B Schmidt-Weber[5], Nikola Müller[4, ‡], Stefanie Eyerich[5, ‡], Kilian Eyerich[1,2,3, *,‡]

Affiliations:

[1]Division of Dermatology and Venereology, Department of Medicine Solna, and Center for molecular medicine, Karolinska Institutet; Stockholm, Sweden;

[2]Unit of Dermatology, Karolinska University hospital, Stockholm, Sweden;

[3]Department of Dermatology and Allergy, Technical University of Munich, Munich, Germany;

[4]Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany;

[5]ZAUM – Center of Allergy and Environment, Helmholtz Center and Technical University of Munich, Munich, Germany;

[6]Department of Dermatology, University of Michigan, Ann Arbor, MI, USA;

[7]Department of Mathematics, Technical University of Munich, Garching, Germany;

[8]Institute for Computational Biomedicine, Englander Institute for Precision Medicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10021, USA

*Correspondence to: rib4003@med.cornell.edu or kilian.eyerich@ki.se

†N.G.Z., R.B., F.L. and M.J. contributed equally to this work.

‡N.M., S.E. and K.E. contributed equally to this work.

Abstract: Chronic inflammatory diseases of the cardiovascular system, brain, gut, joints, skin and lung are characterized by complex interactions between genetic predisposition and tissue-specific immune responses. This heterogeneity complicates diagnoses and the ability to exploit omics approaches to improve disease management, develop more effective therapeutics, and apply precision medicine. Using skin inflammation as a model, we developed a bio-computational approach that assigns deep clinical phenotyping information to transcriptome data of lesional and non-lesional skin (564 samples) to identify biologically-relevant gene signatures. This identified previously unknown key factors, including CCAAT Enhancer-Binding Protein Beta (CEBPB) in neutrophil invasion, and Pituitary Tumor-Transforming 2 (PTTG2) in the pathogenic epithelial response to inflammation. These were validated using genetically-modified human skin equivalents, migration assays, and *in situ* imaging. Thus, by combining deep clinical phenotyping and omics data with sophisticated bio-computational algorithms we present a methodological advance to identify hidden drivers of clinically-relevant biological processes within omics datasets.

One Sentence Summary: Deciphering the pathogenesis of chronic inflammatory diseases by assigning transcriptome profiles to deep clinical phenotyping.

Main text:

Applying computing power to data with high volume and variety, so-called "big data", plays a pivotal role in biomedicine (*1*). For example, deep neural networks can classify diseases from images at least as well, if not better, than specialists, as demonstrated for malignant melanoma (*2*), cardiology (*3*) and breast cancer (*4*). Machine learning approaches are also used to guide diagnoses from laboratory results (*5*), and whole exome sequencing data sets can be used to identify driver mutations in monogenetic diseases and cancer (*6*). In melanoma and glioblastoma, deep neural networks can even be used to design personalized vaccination strategies for patients with impressive clinical success (*7, 8*). However, not all disease conditions are equally amenable to these approaches. For example, identifying causative factors in chronic inflammatory processes is challenging as they are clinically heterogeneous and characterized by a complex interplay between genetic predisposition, immune responses against environmental-, self-, or microbiome-derived antigens, and tissue-specific immune response patterns of e.g. the gastrointestinal system, joints, lung, or skin. In addition and contrast to cancer, which is usually characterized by mutation-driven inexorable progression, inflammatory diseases typically show a relapsing-remitting disease course with a broad spectrum of phenotypes (*9*) ultimately resulting in a pronounced data variety wherein the various distinct signals are not well separated. In line with this, standardized experimental and animal models of human inflammation do not sufficiently reflect the comprehensive human situation (*10-12*), and available human big data sets are insufficiently characterized to dissect disease heterogeneity or identify distinct disease endotypes.

As they are relatively easy to assess and phenotype, non-communicable inflammatory skin diseases (ncISD) represent prototype diseases for chronic inflamation to develop a strategy that identifies relevant molecular events within a complex network of overlaying gene expression signatures. ncISD are a group of several hundred diseases including psoriasis and atopic dermatitis (AD). They are defined by physicians using a subjective "best fit" model when investigating a variety of attributes such as clinical phenotype, tissue and laboratory analysis (*13*). Today, three umbrella phenotypes can be defined according to the immune response pattern in the skin: type 1 ncISD, characterized by cytotoxic immune cells in the skin (so-called interface dermatitis) and comprising the diseases lichen planus and cutaneous lupus; type 2 ncISD, characterized by impaired epidermal barrier and innate immunity and comprising eczematous diseases (atopic dermatitis, dishydrotic eczema, nummular eczema, hyperkeratotic-rhagadiform eczema); and type 3 ncISD defined by Th17 immunity and the presence of neutrophil granulocytes in the skin, e.g. psoriasis and variants (guttate psoriasis, pustular psoriasis) as well as Pityriasis rubra pilaris. Other diagnoses such as cutaneous lymphoma or the pre-malignancy parapsoriasis as well as cutaneous side effects to biologics can be more variable and are not generally attributed to one of these three ncISD clusters. The subjective nature of diagnosing together with the heterogeneity and dynamics of ncISD result in overlapping phenotypes and potential misdiagnoses. This problem is general to many human complex inflammatory diseases of the cardiovascular system, brain, gut, joints, skin or lung, which are increasing in prevalence and account for the majority of global years lived with disability as well as deaths (*14*). Patient's disease burden and socio-economic costs are high and will dramatically increase (*15*). It is estimated that nearly one trillion dollars are wasted due to inefficient treatments alone in the US each year (*16*), so action towards precision medicine is urgently needed. Just like in many chronic inflammatory diseases, there is a clear progress in the development of novel therapeutic agents for the two most common ncISD, psoriasis and Atopic Dermatitis (*17*), but even highly specific biologic therapies still have a substantial number of non-responders (*18, 19*), and causative therapies are missing. Meanwhile, the ability of scientific methods to assess a multitude of biologic processes in an unlimited number of tissue samples is rising (*20-24*). Given that the costs can be reduced by technological progress, the major limitation of research in the field of complex diseases is the

ideal approach of bio-computational interpretation – and the assigned metadata, in particular clinical information at the time of tissue collection.

Thus, to achieve a substantial breakthrough on the path towards precision medicine, the complex pathogenesis of inflammatory diseases such as ncISD and the resulting disease heterogeneity need to be acknowledged in research strategies. Here, we applied deep clinical phenotyping to 287 ncISD patients using 62 attributes including clinical picture, skin architecture, laboratory parameters, and information on the patient´s history, and generated RNA-seq transcriptomes of tissue biopsies of paired lesional and non-lesional skin. To trace gene expression signatures unique to underlying biological processes as captured by the attributes of the deep clinical phenotypes, we developed a novel bio-computational framework, called AUGER. AUGER created a gene expression atlas of chronic cutaneous inflammation to an unprecedented granularity level. Thus, we could identify previously unknown genes driving relevant biological processes. Experimental validation of newly identified key factors shed light on the drivers of neutrophil granulocyte migration into tissue and the pathogenic epithelial response to inflammation that ultimately modulate skin tissue architecture.

Results:

Deep phenotyping explains substantially more transcriptomic variance than the diagnosis alone

Given the heterogeneity in ncISD patients, we hypothesized that current clinical diagnoses only partially reflect the molecular events occurring in the skin. To test this, we collected lesional skin biopsies from 287 patients covering 13 different diagnoses of ncISD (Fig. 1A) and performed RNA-sequencing (RNAseq). Principal component analysis (PCA) of the normalized read count of the RNAseq data showed a mixing of ncISD type 1, type 2, and type 3 diagnoses in the first two major principle components (PC1 and PC2; Fig.1A), which indicates a discrepancy between the diagnosis and the similarity of the transcriptomes at least in some patients. To exclude the possibility that transcriptional heterogeneity is driven by intra-individual differences, we included normalized read counts of RNAseq measurements of autologous non-lesional skin biopsies (n = 564 samples in total) and performed again a PCA on the top-most variable genes (Fig. S1). The status of the skin (lesional or non-lesional) explained the separation of PC1 and in total 32 % of the variance in the gene expression data set, with genes related to skin inflammation driving the separation (Fig. S1A, B). Another 9 % of the variance mainly in lesional samples was explained by gender, which was confirmed by the expression of genes located on sex chromosomes (Fig. S1A, B). Thus, we performed unsupervised hierarchical clustering of lesional transcriptomes excluding genes encoded on the sex chromosomes. There were still patients with different diagnoses in the same clusters (Figure S1C), indicating that gender correction alone is insufficient to improve the correlation of transcriptomes and clinical phenotypes.

To identify how other disease-relevant factors impact the transcriptional heterogeneity seen in patients, we performed deep phenotyping of the 287 patients using 62 attributes from the data recorded by dermatologists during diagnosis that reflect clinical picture, histologic architecture, laboratory results, investigation of comorbidities, and personal and family history (Fig. 1B, Fig. S2). This generated 287 unique patient fingerprints (Fig. S3). We first analyzed the fingerprints from four patients with the diagnosis psoriasis (n = 2) or AD (n = 2). Three of these clustered closely based on transcriptional profiles, while one AD patient was located far from the other three patients in the PCA analysis (Fig. 1C). At this single patient level, patients with transcriptionally-similar profiles indeed represented individuals whose attribute-based fingerprints showed more similarities despite the fact they had different diagnoses than patients with the same diagnosis but different locations in the PCA i.e., transcriptionally dissimilar profiles (Fig. 1C). Hence, we investigated whether deep clinical phenotyping by attributes is superior over the physicians' diagnosis to explain variance in the transcriptome. A linear mixed model was used to attribute variance to either diagnosis or separate attributes in lesional data. We found that patients' diagnoses only explained 7.8 % of the total transcriptional variance in the skin (Figure 1D). In contrast, the 62 attributes of the deep phenotyping together explained 36.7 % of the total transcriptional variance (Fig. 1D, E). Thus, multiple, clinically-relevant attributes in a patient more closely drive their lesional transcriptional profiling than the clinician's diagnosis alone, and may improve disease management and enable the identification of novel, biologically- and clinically-relevant molecular events.

Unique attribute-gene signatures identified by effective demixing of the transcriptional signal

Since single attributes are not specific for a distinct disease but still contribute to shaping the lesional skin transcriptome, the traditional approach of comparing diseases is limited due to a high background noise caused by these conserved attributes. Thus, we developed a bio-computational algorithm that dissects the lesional skin transcriptome into gene expression signatures of each attribute to identify previously unknown genes driving relevant biological processes - AuGER (**A**ttrib**u**te **G**ene **E**xpression **R**egularization, Fig. 2A). First, to simplify

the data by removing redundant attributes, we performed pairwise correlation analysis, which revealed groups of highly correlated attributes (Fig. 2A 1-3, Fig. 2B, Table S3). As expected, attributes used to describe an ncISD disease type highly correlated with each other: specifically, the ncISD type 1 attributes, "number of dyskeratoses", "interface dermatitis qualitative" and "quantitative", "localization of dyskeratoses"; the ncISD type 2 attributes, "serum crusts", and "spongiosis"; and the ncISD type 3 attributes, "neutrophils qualitative" and "quantitative", and "parakeratosis" (Figure 2B). One representative attribute (= core attribute) per group of highly correlated attributes was selected based on clinical relevance, favouring quantitative measures over qualitative discretization, resulting in a total of 24 core attributes used for further analysis (Fig. 2C).

To identify unique gene sets whose transcriptional activity is associated with core attributes, multivariate regression with elastic net penalty was performed. For each of the 24 core attributes regression analysis was carried out, whereat the respective core attribute was used as the response variable and the genes of the complete transcriptome data as the predictor variables. The remaining 23 core attributes were used as covariates to control for their impact on the response variable. For each gene, a regression coefficient was calculated for the respective attribute, and all genes with non-zero regression coefficients were collected (Fig. 2A 4-5). To finally identify the disease-specific associations between individual attributes and genes that have clinical relevance for ncISD, we not only analysed individual attribute signatures linked to the corresponding lesional transcriptome (n = 287) but also to the respective autologous non-lesional transcriptome obtained (n = 277). To filter out attribute-associated genes that are not specific for inflammation, only genes with non-zero coefficient in lesional but not in nonlesional skin were selected (Fig. 2A 6). To improve robustness, regression analysis for both lesional and non lesional transcriptome was conducted five times, respectively, using five sets of imputed core attributes. Specific unique and shared associations were obtained between a total of 621 genes and 18 of the 24 attributes creating a bipartite network (Fig. 2A 7, Figure S4). For the first time, AuGER acknowledges disease heterogeneity of ncISD by dissecting diseases into multiple clinically-relevant attributes and associating these attributes to gene signatures. These gene signatures reveal the molecular basis of complex and biologically-relevant processes, e.g. the presence of neutrophil granulocytes in the skin or pathogenic epithelial thickening as a consequence of inflammation, so-called acanthosis.

The gene expression landscape of ncISD

To understand how the attributes are connected and if their interaction reflects ncISD types, we created an atlas of skin inflammation with its underlying gene signature by compiling a network of gene and attribute nodes linked by an edge when the attribute-gene pair was identified by AuGER (Fig. 2C, Figure S5, Table S3). On the one hand, this atlas reflects core attributes that represent hallmarks of the three ncISD clusters, type 1 (lichen planus and cutaneous lupus, here identified by the core attribute "interface dermatitis"), type 2 (eczema variants which are typically associated to "dryness" and "eosinophils"), and type 3 (psoriasis variants represented by the core attribute "neutrophils").

On the other hand, the atlas also illustrates attributes which are not typically associated with one of the three ncISD clusters, but still have an impact on the skin transcriptome. These include scalp or nail involvement, clinical presentation such as scaling, papules, or erythema; or histo-pathological presentation such as acanthosis or orthokeratosis (regular formation of the upper epidermis). These attributes may occur in several diseases and thus increase the unspecific background noise in traditional approaches comparing different diseases. Finally, external factors influencing the lesional skin transcriptome are also captured. This is represented by numerous genes associated with the core attributes "alcohol consumption" and "smoking" (Fig. 2C).

Notably, AuGER identified only one gene associated with the attribute "pruritus", namely the T cell recruiting chemokine *CCL18* (*25*). *CCL18* was shared with the core attribute "eosinophils" and thus connected to type 2 ncISD, namely eczema. This is in line with the fact that most eczema variants are highly pruritic (*26*). Indeed, levels of *CCL18* were highest in eczema variants (AD and nummular eczema) as compared to psoriasis (Fig. S6A). This was confirmed in an independent data set consisting of skin from healthy volunteers (n = 38) and from patients suffering from AD (n = 27) or psoriasis (n = 37). Here, expression of CCL18 was again significantly higher in AD than in the healthy controls (log2 fold change: 4.9, p $1.4*10^{-13}$) or psoriasis (3.1, p $1.4*10^{-7}$) (Fig. S6B). In a third cohort, CCL18 expression was higher in lichen planus lesional skin (n = 20) as compared to healthy controls (n = 38). Just like eczema, lichen planus is typically pruritic (*27*)(Fig. S6B).

All attributes except "mucin" shared at least one link with another attribute in the network, thus constructing a comprehensive gene map of skin inflammation. "Mucin" is a core attribute comprising an enhanced level of extracellular material (*28*). It is not associated with an inflammation cluster and is usually the result of inflammation of deeper skin layers rather than of the epidermis. Connections of the individual gene signatures of each attribute were also identified at the level of pathway analysis. Some pathways were associated with only one attribute, e.g. the expected cytotoxicity-mediating pathway "IFN-gamma signaling" associated with the ncISD type 1 core attribute "interface dermatitis" (*29*). Numerous pathways were associated with at least five different attributes. For example, pathways of cell cycle and proliferation, such as "EGF signaling", "Ras signaling", "Oncostatin M signaling", and the stem cell factor "Wnt signaling", connected the clinical attributes "elevation" (palpable skin lesions) and "papules", both of which may be related to exaggerated proliferation of the epidermis (Fig. S5). Next, the network was developed further by considering protein-protein interactions. Taken together, the atlas of skin inflammation reflects interactions of biologically-relevant processes, and furthermore gives a much higher resolution of genes driving these processes due to acknowledging the heterogeneity of ncISD through AuGER.

Novel insights into neutrophil granulocyte biology in tissue

The comprehensive atlas of the transcriptome landscape of ncISD delivers a plethora of previously unknown connections of biologic processes with expression of genes. To understand whether these genes are pathogenic, we validated genes associated to the attribute "neutrophils", which was defined as the presence of neutrophil granulocytes in lesional skin as observed by dermatopathology. AuGER identified an association of 35 genes with this attribute (Fig. 3A). While some of these genes such as *CXCL1* or *CXCL8* are well-known drivers of neutrophil migration and/or biology (*30*), others were previously unknown in this context. As expected, the ncISD type 3 core attribute "neutrophils" was most prominent in psoriasis patients, however, intra- and inter-disease variability was high (Fig. 3B, C). Pathway enrichment of the identified genes showed they could be attributed to inflammation and general processes such as protein binding (Fig. 3D). For biological validation of the attribute-gene association, we focused on five genes *CEBPB*, *HCAR3*, SHC1, *HS6ST1* and *SLC23A2* as they were centrally located in the network (Fig. 2C) and previously not linked to skin diseases. We first analyzed the effect of stimulation with pro-inflammatory cytokines IL-17A, TNF-α or LPS on gene expression in two potential target cells, primary human keratinocytes and neutrophil-like HL-60 cells. Stimulation induced gene expression only of *CEBPB* consistently in both cell types as compared to unstimulated controls (Fig. 4A).

C/EBP factors, to which CEBPB belongs, have been found to regulate cytokine expression in human neutrophils (*31*). To gain more insight into the function of *CEBPB* in neutrophil biology in particular into the interaction of neutrophils and keratinocytes, we specifically knocked-out *CEBPB* and the irrelevant control gene *PTTG2* in keratinocytes using

CRISPR/Cas9 technology. *CEBPB*, but not *PTTG2* knock outs, resulted in significant reduction of central chemokines for neutrophil migration and activation (CXCL8, CXCL1, CXCL2, CXCL5, IL-6) in the supernatant of keratinocytes stimulated with IL-17A (Fig. 4B). Meanwhile, *CEBPB* knock-out cells produced significantly higher amounts of the growth factor GM-CSF and the chemokines CCL5 and CCL22, indicating that loss of CEBPB does not hamper the ability of keratinocytes to promote bone marrow neutrophil maturation, or migration of eosinophils, macrophages or dendritic cells (Fig. 4B). To validate that CEBPB regulates neutrophil migration to the skin, we quantified migration of primary human neutrophils towards the supernatant of *CEBPB* or *PTTG2* knock-out keratinocytes, compared to the supernatant of wild type controls. Here, knock-out of *CEBPB* significantly reduced the capacity of keratinocytes to regulate neutrophil migration (Fig. 4C). Furthermore, knock-out of *CEBPB* directly in the neutrophil-like HL-60 cells resulted in diminished migration towards both CXCL8 and keratinocyte supernatant (Fig. 4D). Thus, CEBPB is a pivotal regulator of neutrophil biology in the skin, acting at both the keratinocyte and the neutrophil level.

Pathogenic epidermal hyperplasia is driven by previously unknown genes

Pathogenic hyperplasia of the epidermis caused by excessive proliferation and disturbed differentiation of keratinocytes, so-called acanthosis, is a process frequently observed in different kinds of skin inflammation. Therefore, acanthosis is a representative example of how AuGER filters and identifies previously unknown biologic drivers of this conserved clinical phenotype. 69 genes were associated with acanthosis (Fig. 2C, Fig. 5A). Acanthosis is neither specific for a ncISD type, nor is it observed in all patients suffering from an ncISD. In line with this, in our cohort, acanthosis as an attribute was variably distributed across the 13 types of ncISD (Fig. 5B) and unsupervised hierarchical clustering of genes associated with acanthosis in all 287 patients did not correlate with distinct diseases (Fig. 5C). These acanthosis-associated genes were assigned to pathways describing cell-cell adhesion and activity (Fig. 5D).

For biological validation of the attribute-gene association, we selected *PTTG2*, *SYNE1*, and *CEBPB*. Keratinocytes stimulated with the pro-inflammatory cytokines IL-17A and TNF-α showed an upregulation of *CEBPB* and a downregulation of *SYNE1* gene expression, while *PTTG2* expression remained unaltered (Fig. 6A). Knock-out of either *PTTG2*, *SYNE1*, or *CEBPB* in primary human keratinocytes did not alter the cellular morphology or the architecture of three-dimensional human skin equivalents (Fig. 6B). The cytokine IL-22 is a known inducer of keratinocyte growth and induces acanthosis (*32*). We next investigated the effects of *CEBPB*, *SYNE1* and *PTTG2* knock-out on epidermal thickness as a proxy for acanthosis in response to IL-22. In line with the differences in gene regulation, knock-out of both *PTTG2* and *CEBPB* significantly inhibited the induction of acanthosis by IL-22 compared to the control (noRNP) model, with *CEBPB* preventing it completely. In contrast, knock-out of *SYNE1* markedly increased the thickness of the three-dimensional skin model (Fig. 6B, C; Fig. S8). Notably, proliferation of keratinocytes was also reduced by knock-out of *CEBPB* or *PTTG2* but induced by *SYNE1* (Fig. 6C, D, E). Finally, knock-out of *CEBPB* resulted in altered epithelial differentiation after IL-22 stimulation, namely in clearly diminished expression of the early differentiation marker keratin 10, and keratin 16, which is typically associated with proliferative keratinocytes (*33*) (Fig. 6E). Hence, CEBPB and PTTG2 are essential for the development of acanthosis, while SYNE1 is a negative regulator of epidermal thickness.

Discussion:

Here, we show that previously unknown biologic drivers of key pathogenic events in human inflammatory diseases can be identified from transcriptome information when disease heterogeneity is acknowledged. We investigate non-communicable inflammatory skin diseases (ncISD) as prototype examples of heterogeneous human inflammatory disorders.

We observed that assigning disease names to data sets alone only explains a minority of the transcriptome variation. In contrast, metadata with a higher granularity of the observed clinical heterogeneity explain a substantially higher proportion of the transcriptome variability associated with inflammation. Furthermore, we demonstrate that common hallmarks of the interplay between skin epithelia and immune cells are shared between different patient cohorts independent of the disease background. We and others recently followed this approach for ncISD that share the histological hallmark “interface dermatitis”, which represents cytotoxic activity of lymphocytes at the basal membrane of the epidermis and is frequently observed in type 1 ncISD such as lichen planus and lupus (*29, 34*). Here, identification of the molecular mechanism, type 1 immune mediated epithelial apoptosis and necroptosis, led to potential novel therapeutic strategies (*35*). Our current analysis now takes into account other dependent clinical attributes and gives an even clearer picture of the genes involved in interface dermatitis.

Even in settings that are heavily researched and where numerous drivers are known, such as neutrophil biology, assigning more granular metadata of clinical phenotyping identifies previously unknown, but important mechanistic factors. In particular, our newly developed algorithm AuGER identified CEBPB as a master regulator of neutrophil biology in the skin. C/EBP proteins are known to be involved in cytokine production such as CXCL8 and CCL3 by neutrophils (*31, 36*). We show here that *CEBPB* is also expressed in keratinocytes, and knock-out of *CEBPB* in primary human keratinocytes results in reduced chemokine secretion and reduced neutrophil migration. This newly identified dual mechanism of action to inhibit neutrophil inflammation in tissue renders CEBPB a potentially powerful therapeutic target for neutrophilic inflammatory diseases of the skin, such as psoriasis or pyoderma gangrenosum but also for diseases beyond the skin as for example reactive arthritis or neutrophil-rich chronic obstructive lung disease (COPD).

Our approach not only reveals the most significant genes contributing to an attribute, but *vice versa* also identifies the mechanistic significance of single genes for corresponding attributes representing complex mechanistic processes. This was proven for the less restricted attribute “acanthosis”. Acanthosis corresponds to reactive thickening of the epidermis in inflammatory conditions and it can be observed in multiple ncISD but is not part of a disease definition (*37*). *PTTG2*, a hub gene for acanthosis in our analysis, has not been proposed in the context of acanthosis before. PTTG2 is involved in cell viability and migration by regulating the expression of vimentin and E-cadherin and overexpression of *PTTG2* was described in psoriatic epidermis (*38, 39*). We now show that PTTG2 is directly involved in acanthosis, as inhibiting this single factor significantly impaired acanthosis formation *in vitro*. *SYNE1*, another gene significantly correlated to acanthosis from our network, encodes an important scaffold protein. In genetically modified mice, loss of *SYNE1* leads to thickening of the epidermis (*40*), an observation we validated in the human system by showing increased thickness of three-dimensional skin equivalents comprised of *SYNE1* knocked-out keratinocytes. Interestingly, *CEBPB*, associated with the attribute “neutrophil” was also associated with “acanthosis” and the third candidate gene we validated for its contribution to acanthosis *in vitro*. In fact, human skin equivalents of keratinocytes lacking *CEBPB* were vital and did not show any morphological alterations, but they did not respond to IL-22, an inflammatory stimuli that normally induces acanthosis (*32*). Hence, CEBPB seems an ideal target to both address diseases in which neutrophils and acanthosis are concomitantly present, in particular in psoriasis, hidradenitis suppurativa, or severe acne.

Our *proof-of-concept* study has limitations. First, phenotyping was partially based on subjective information, such as the patient´s narrative, clinical examination and histological evaluation. Second, though we performed RNA-Sequencing of 287 patient's tissue samples, the sample size is still limited, and underestimation of real biologic effects cannot be ruled out. To raise our network approach to the next level and with it eventually identify robust predictive markers for medically-relevant outcomes such as response to therapy, more data sets from multiple sources need to be fed into the network. However, a highly advanced neural network of diseases can only be generated provided that each molecular data point is assigned to metadata as it has been done in this study. However, the study clearly shows already at this stage that assigning clinical phenotypes to large datasets critically determines the quality of outcome of the analysis. Phenotyping methods such as image-based machine-learning will further improve our novel approach in the future.

Taken together, exemplifying ncISD we show here that combining deep clinical phenotyping with transcriptome datasets and novel bio-computational algorithms might be key to define pathogenic events in complex processes such as chronic human inflammation and thus build the basis for precision medicine in the field.

References:

1. I. Rahwan *et al.*, Machine behaviour. *Nature* **568**, 477-486 (2019).
2. A. Esteva *et al.*, Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115-118 (2017).
3. D. Dey, A. Lin, Machine-Learning CT-FFR and Extensive Coronary Calcium: Overcoming the Achilles Heel of Coronary Computed Tomography Angiography. *JACC Cardiovasc Imaging*, (2019).
4. S. M. McKinney *et al.*, International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89-94 (2020).
5. S. Cristiano *et al.*, Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389 (2019).
6. N. K. Hayward *et al.*, Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175-180 (2017).
7. P. A. Ott *et al.*, An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217-221 (2017).
8. N. Hilf *et al.*, Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* **565**, 240-245 (2019).
9. C. Nathan, A. Ding, Nonresolving inflammation. *Cell* **140**, 871-882 (2010).
10. K. Eyerich *et al.*, Human and computational models of atopic dermatitis: A review and perspectives by an expert panel of the International Eczema Council. *J Allergy Clin Immunol* **143**, 36-45 (2019).
11. J. E. Hawkes, J. A. Adalsteinsson, J. E. Gudjonsson, N. L. Ward, Research Techniques Made Simple: Murine Models of Human Psoriasis. *J Invest Dermatol* **138**, e1-e8 (2018).
12. N. Choudhary, L. K. Bhatt, K. S. Prabhavalkar, Experimental animal models for rheumatoid arthritis. *Immunopharmacol Immunotoxicol* **40**, 193-200 (2018).
13. K. Eyerich, S. Eyerich, Immune response patterns in non-communicable inflammatory skin diseases. *J Eur Acad Dermatol Venereol* **32**, 692-703 (2018).
14. G. B. D. Disease, I. Injury, C. Prevalence, Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789-1858 (2018).
15. N. Global Burden of Disease Health Financing Collaborator, Future and potential spending on health 2015-40: development assistance for health, and government, prepaid private, and out-of-pocket health spending in 184 countries. *Lancet* **389**, 2005-2030 (2017).
16. W. H. Shrank, T. L. Rogstad, N. Parekh, Waste in the US Health Care System: Estimated Costs and Potential for Savings. *JAMA*, (2019).
17. S. Eyerich, M. Metz, A. Bossios, K. Eyerich, New biological treatments for asthma and skin allergies. *Allergy*, (2019).
18. K. Reich *et al.*, Guselkumab versus secukinumab for the treatment of moderate-to-severe psoriasis (ECLIPSE): results from a phase 3, randomised controlled trial. *Lancet* **394**, 831-839 (2019).
19. A. Blauvelt *et al.*, Long-term management of moderate-to-severe atopic dermatitis with dupilumab and concomitant topical corticosteroids (LIBERTY AD CHRONOS): a 1-year, randomised, double-blinded, placebo-controlled, phase 3 trial. *Lancet* **389**, 2287-2303 (2017).
20. D. Aran, Z. Hu, A. J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220 (2017).
21. S. Choobdar *et al.*, Assessment of network module identification across complex diseases. *Nat Methods* **16**, 843-852 (2019).

22. Y. S. Lee *et al.*, Interpretation of an individual functional genomics experiment guided by massive public data. *Nat Methods* **15**, 1049-1052 (2018).
23. H. W. L. Koh *et al.*, iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. *NPJ Syst Biol Appl* **5**, 22 (2019).
24. J. Arloth *et al.*, DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. *PLoS computational biology* **16**, e1007616 (2020).
25. C. Gunther *et al.*, CCL18 is expressed in atopic dermatitis and mediates skin homing of human memory T cells. *J Immunol* **174**, 1723-1728 (2005).
26. S. Weidinger, L. A. Beck, T. Bieber, K. Kabashima, A. D. Irvine, Atopic dermatitis. *Nat Rev Dis Primers* **4**, 1 (2018).
27. S. Steinke *et al.*, Humanistic burden of chronic pruritus in patients with inflammatory dermatoses: Results of the European Academy of Dermatology and Venereology Network on Assessment of Severity and Burden of Pruritus (PruNet) cross-sectional trial. *J Am Acad Dermatol* **79**, 457-463 e455 (2018).
28. A. Fernandez-Flores, M. Saeb-Lima, Mucin as a diagnostic clue in dermatopathology. *J Cutan Pathol* **43**, 1005-1016 (2016).
29. F. Lauffer *et al.*, Type I Immune Response Induces Keratinocyte Necroptosis and Is Associated with Interface Dermatitis. *J Invest Dermatol* **138**, 1785-1794 (2018).
30. F. Lauffer *et al.*, IL-17C amplifies epithelial inflammation in human psoriasis and atopic eczema. *J Eur Acad Dermatol Venereol* **34**, 800-809 (2020).
31. A. Cloutier *et al.*, Inflammatory cytokine production by human neutrophils involves C/EBP transcription factors. *J Immunol* **182**, 563-571 (2009).
32. Y. Zheng *et al.*, Interleukin-22, a T(H)17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature* **445**, 648-651 (2007).
33. R. Moll, M. Divo, L. Langbein, The human keratins: biology and pathology. *Histochem Cell Biol* **129**, 705-733 (2008).
34. S. Shao *et al.*, IFN-gamma enhances cell-mediated cytotoxicity against keratinocytes via JAK2/STAT1 in lichen planus. *Science translational medicine* **11**, (2019).
35. W. Damsky *et al.*, Treatment of severe lichen planus with the JAK inhibitor tofacitinib. *J Allergy Clin Immunol*, (2020).
36. T. Z. Mayer *et al.*, The p38-MSK1 signaling cascade influences cytokine production through CREB and C/EBP factors in human neutrophils. *J Immunol* **191**, 4299-4307 (2013).
37. A. Fernandez-Flores, Lesions With an Epidermal Hyperplastic Pattern: Morphologic Clues in the Differential Diagnosis. *Am J Dermatopathol* **38**, 1-16; quiz 17-19 (2016).
38. X. B. Liu, F. Li, Y. Q. Li, F. Yang, Pituitary tumor transforming gene PTTG2 induces psoriasis by regulating vimentin and E-cadherin expression. *Int J Clin Exp Pathol* **8**, 10887-10893 (2015).
39. C. Mendez-Vidal *et al.*, PTTG2 silencing results in induction of epithelial-to-mesenchymal transition and apoptosis. *Cell Death Dis* **4**, e530 (2013).
40. Y. Luke *et al.*, Nesprin-2 Giant (NUANCE) maintains nuclear envelope architecture and composition in skin. *J Cell Sci* **121**, 1887-1898 (2008).
41. E. Rincon, B. L. Rocha-Gregg, S. R. Collins, A map of gene expression in neutrophil-like cell lines. *BMC Genomics* **19**, 573 (2018).

Acknowledgments: We thank Helen Pickersgill from Life Science Editors for editing the manuscript and Jana Sänger as well as Kerstin Weber for excellent technical support. In addition, we thank the NGS core facility at the Helmholtz Center Munich for help with RNA sequencing and the Biobank Biederstein for providing patient samples. **Funding:** Funding sources: This work was supported by an ERC grant (IMCIS, 676858) and the German Research Foundation (EY97/3-2). **Author contributions:** Conceptualization: N.G.S, F.L., S.E., N.S. M, K.E; Methodology: R.B., N.S.M, M.J.; Validation M.J.; Formal Analysis R.B., L.C.T., J.E.G.; Investigation: M.J., ; Resources: N.G.S, F.L., C.P., S.R., A.S., A.B., P.S, B. F., J.T.; Data Curation: R.B.; Writing –Original Draft: F.L., N.G.S, S.E., K.E.; Writing –Review & Editing Visualization: C.B.S-W, T.B., F.T., J.E.G.; Supervision: N.S.M., S.E., K.E.; Project Administration: N.S.M., S.E., K.E.; Funding Acquisition: K.E. **Competing interests:** Authors declare no competing interest in relation to this article.

Supplementary Materials:

Materials and Methods

Figures S1-S7

Tables S1-S3

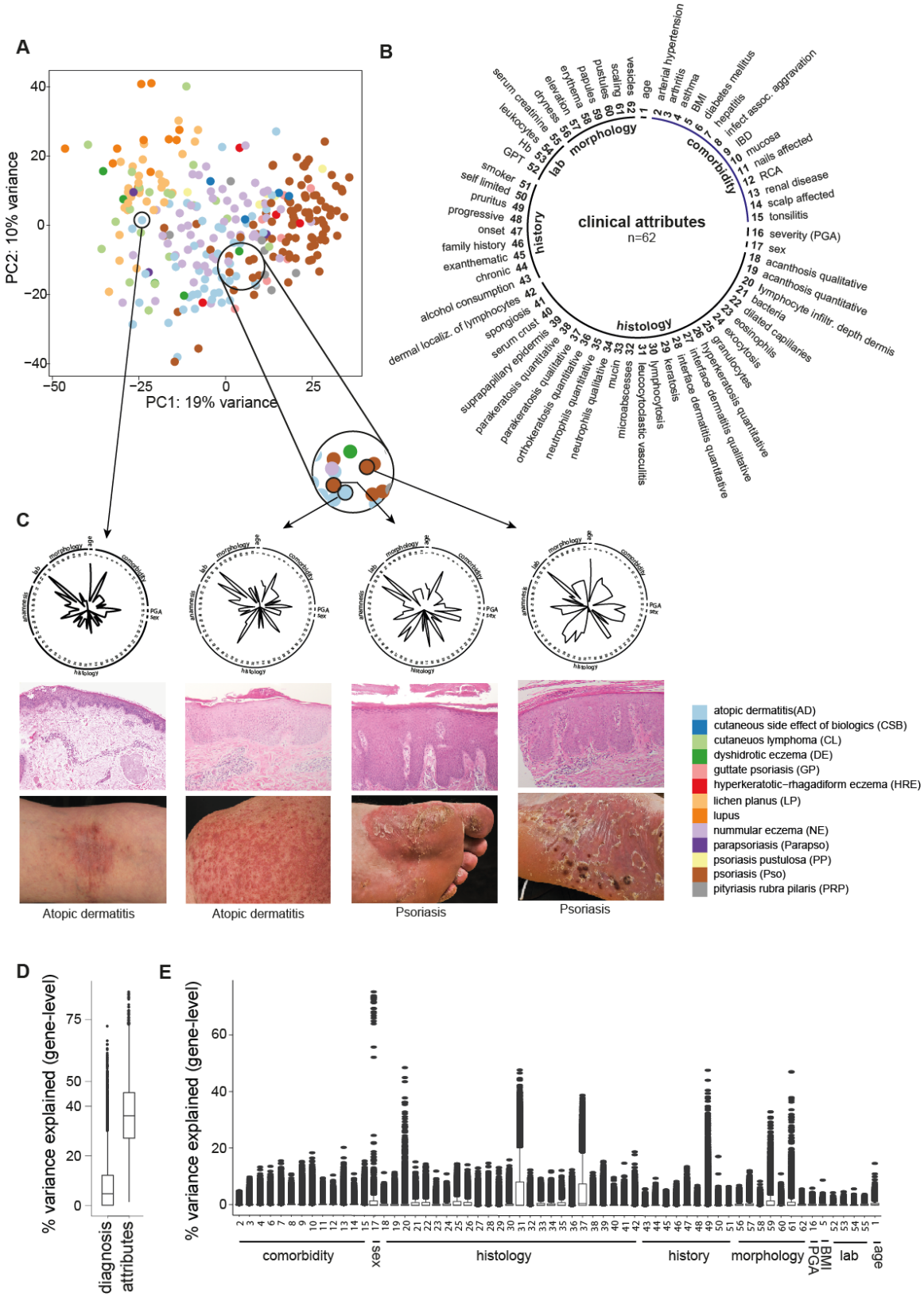

Fig. 1. Transcriptome heterogeneity of ncISD requires deep clinical phenotyping. **A)** Principal component (PC) analysis of gene expression determined by RNA sequencing of lesional skin biopsies obtained from 287 patients suffering from 13 different non-communicable inflammatory skin diseases. Patients are colored by diagnosis as indicated in the legend. The PCA reveals three, albeit not very concise, subclusters consisting of lichen planus/ lupus (type 1 ncISD) (light and dark orange), nummuluar eczema/ atopic eczema (type 2 ncISD) (light purple/light blue) and psoriasis (type 3 ncISD) (brown) **B)** Overview of clinical attributes (n=62) subgrouped into morphology, comorbidity, histology, patient's history (history), and laboratory parameters (lab), used to obtain a deep clinical phenotype of each patient. **C)** Exemplary fingerprints of four patients (two atopic dermatitis and two psoriasis patients) attribute composition (radar plots), histological examination (middle) and clinical picture (right) demonstrating that two patients with a transcriptionally similar profile also have a similar clinical phenotype (radar plot) despite receiving a different clinical diagnosis (patients highlighted in the circle). However, patients with the same diagnosis (e.g. atopic dermatitis as indicated) may have a heterogeneous clinical phenotype combined with transcriptional diversity. **D)** To understand if clinical phenotyping is superior to the diagnosis in explaining the variance in the transcriptome data, a linear mixed model was used to attribute variance to diagnosis and attributes revealing that diagnosis can explain less variance (7.8%) of the data set than attributes (36.7%). **E)** Overview of the contribution of each individual clinical attribute to the variance in the transcriptome data set. BMI: body mass index, GPT: Glutamat-Pyruvat-Transaminase, Hb: hemoglobin, IBD: inflammatory bowel disease, RCA: rhinoconjunctivitis allergica. See also Figure S1, S2 and S3.

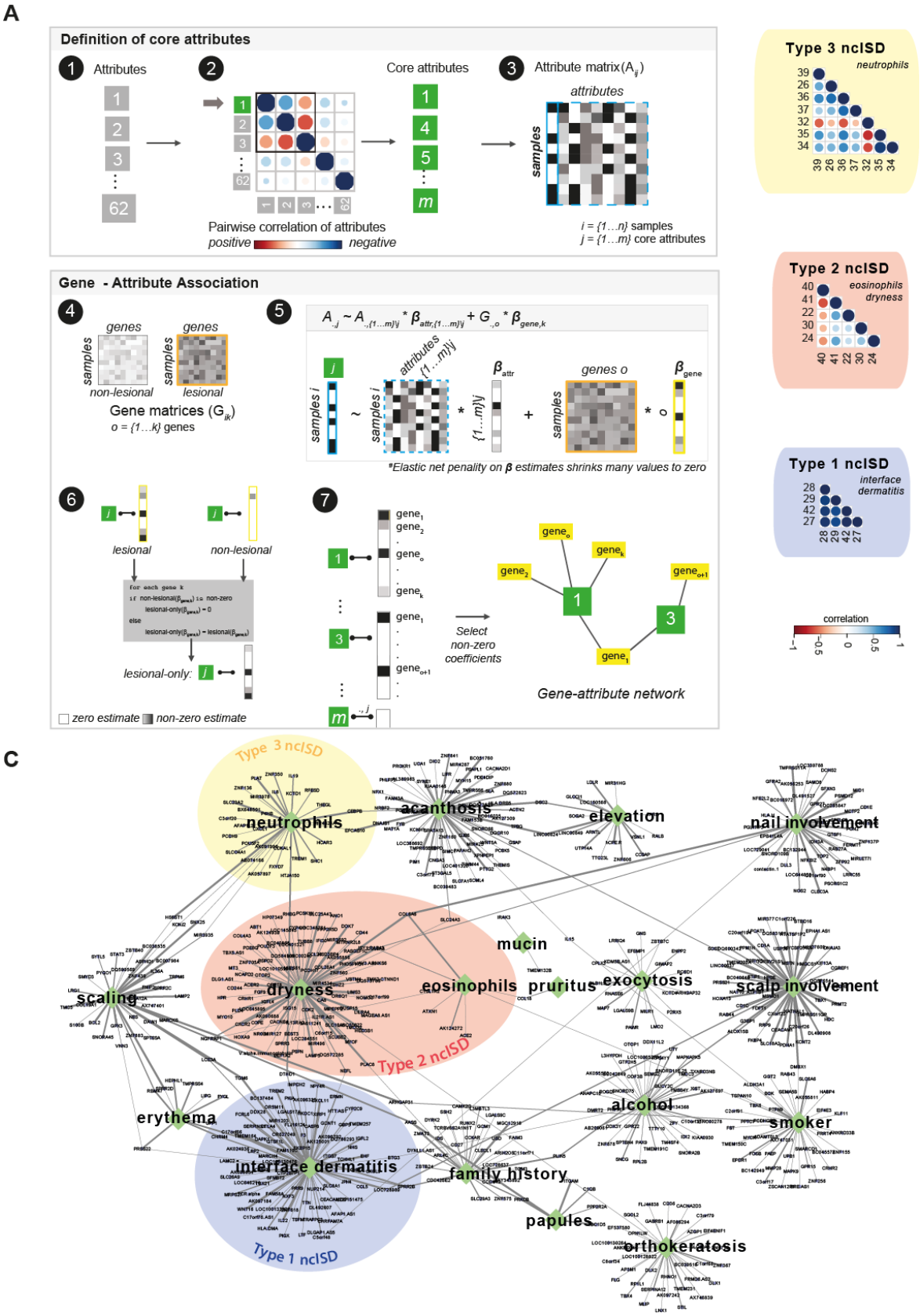

Fig. 2. The concept and resulting network of AuGER (**A**ttrib**u**te **G**ene **E**xpression **R**egularization). **A)** Schematic of the AuGER approach. 1st step: Definition of core attributes. The 62 attributes (1) of all samples *i*= {1…*n*} were subjected to pairwise comparisons (2) to identify highly correlated attributes. Clinically most relevant attributes representative for a set of highly correlated attributes were selected resulting in a set of 24 core attributes *j* (*j*= {1…*m*} (2) and thus in the attribute matrix $A_{i,j}$. (3), 2nd step: Gene-attribute association. First, RNASeq count data was split into lesional and non-lesional sets of *k* genes (*o* = {1… *k*}, gene matrix G_{ik}), respectively) (4). Next, multivariate regression with elastic net penalty was used to identify the association between each core attribute (blue box) and gene expression (here illustrated for lesional skin only, orange box). The m-1 attributes not in response were used as covariates to control for the influence on response variable (dashed blue box). Beta coefficients β_{attr} and β_{gene} (yellow box) for each predictor variable were estimated (5), however, only genes unique to lesional skin were selected (6). Next, genes with a non-zero β_{gene} coefficient in lesional skin only were filtered for each attribute and the association of genes and clinical attributes was visualized as a bipartite network (7). **B)** Exemplary highly correlated attributes belonging to type 1, type 2, or type 3 ncISD clusters with their core attribute name in italics. **C)** Network visualization of gene sets derived from AuGER. Attributes (green diamonds) and genes represent nodes, and thickness of edges represent strength of an association. Colors indicate core attributes of the three ncISD clusters. See also Figure S4, S5, and Table S3.

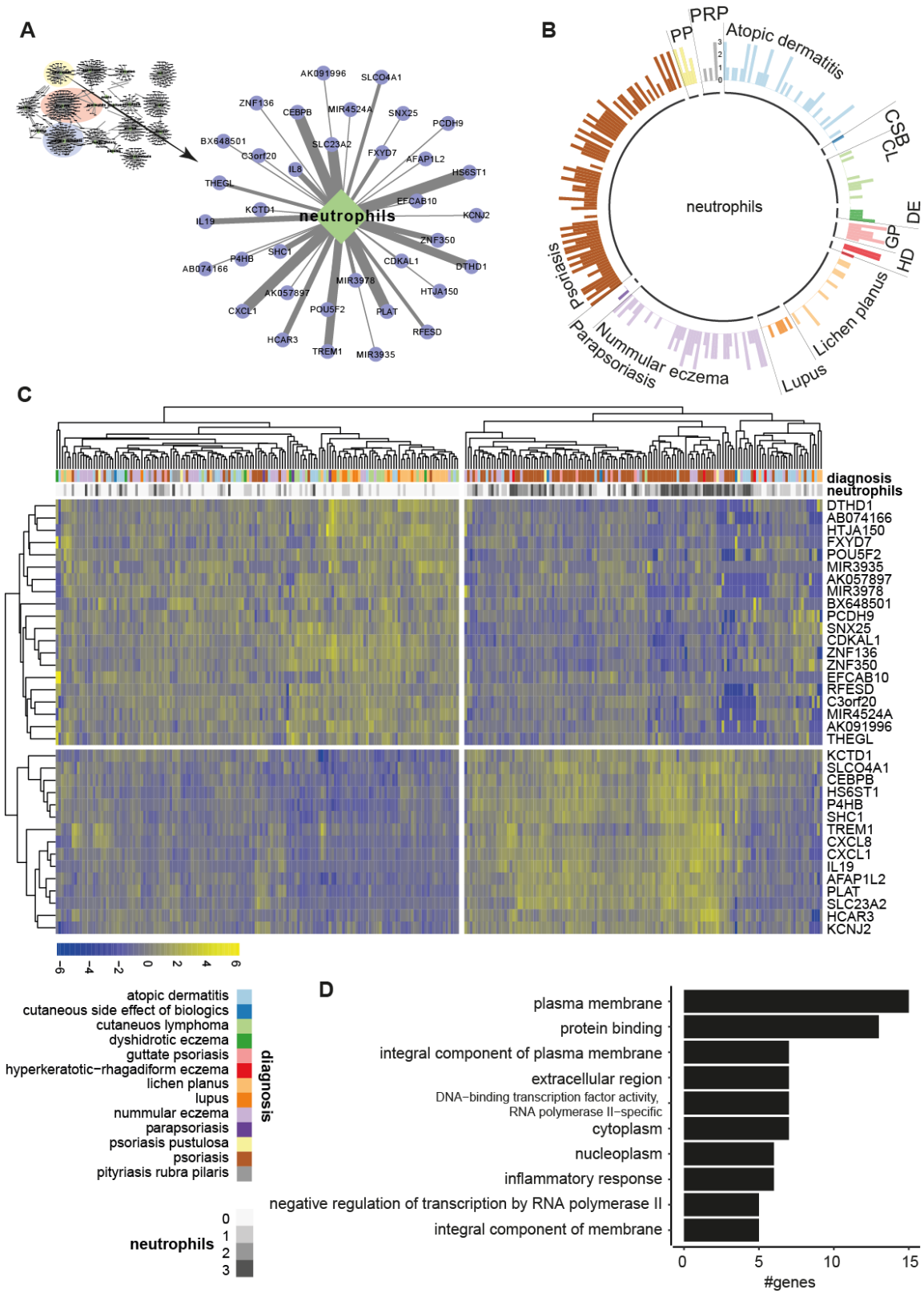

Fig. 3. Gene expression landscape of neutrophil granulocytes in inflamed tissue. **A)** A focus on the clinical attribute 'neutrophils' and its associated genes. Line thickness indicates the strength of gene-attribute association determined by the presence of the association in five iterating bio-computational cycles (thickest line=five times present). **B)** Distribution of the attribute 'neutrophils' across the 13 investigated non-communicable inflammatory skin diseases. Each bar represents one patient, and the height of the bar indicates the number of neutrophils present in skin histology with 0 = no, 1 = low, 2 = moderate, 3 = high numbers of lesional neutrophils (see Fig. S2). **C)** Hierarchical clustering of neutrophil-associated genes assigned with diagnosis (colors) and presence of neutrophils (grayscale). **D)** Gene ontology terms associated with the neutrophil gene set. CL: cutaneous lymphoma; CSB: cutaneous side effect of biologics; DE: dishydrotic eczema; GP: guttate psoriasis; Parapso: parapsoriasis; PP: psoriasis pustulosa; PRP: pityriasis rubra pilaris.

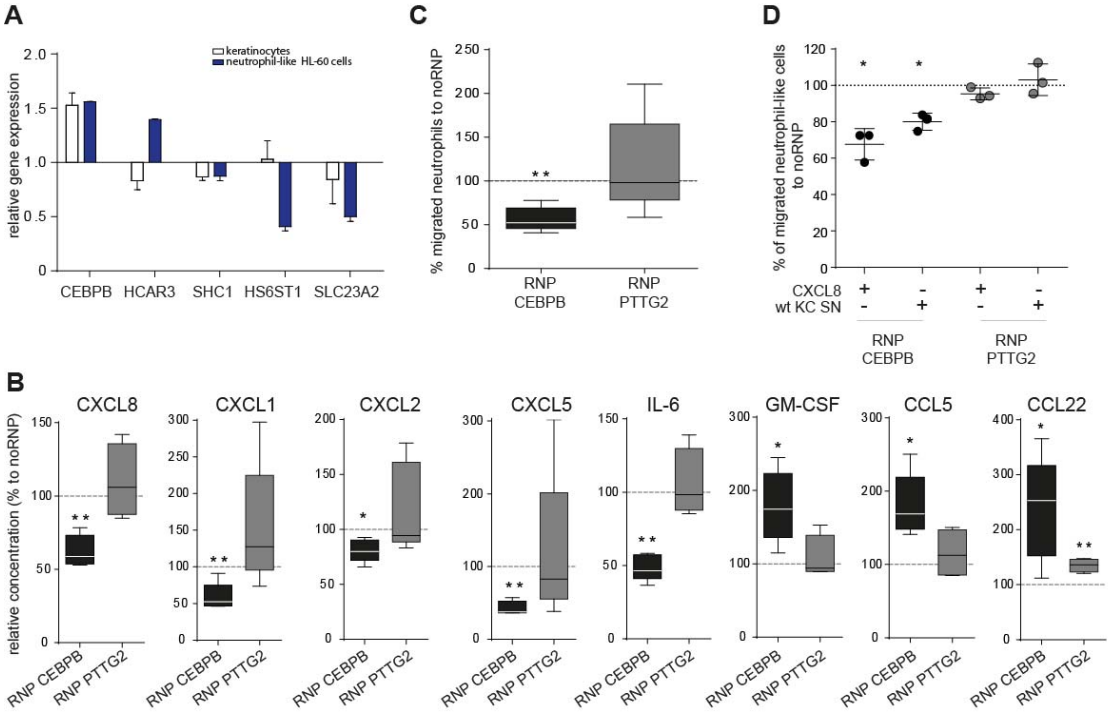

Fig. 4. CEBPB is a key regulator of neutrophil granulocyte biology. **A)** Five genes with a strong association to the attribute 'neutrophils' and a yet unknown function in this field were selected and their relative gene expression was evaluated in primary human keratinocytes (stimulated with IL-17A and TNF-α) and neutrophil-like cells (differentiated HL60 cells) (stimulated with LPS) compared to non-stimulated controls by real time PCR. **B)** *CEBPB* and *PTTG2* were knocked out in primary human keratinocytes using the RNP-CRISPR technology. Knock-out (RNP) and wild-type (no-RNP) keratinocytes were cultured and stimulated with recombinant IL-17A for 72 hours to induce neutrophil-relevant genes. The supernatant was analysed for cytokine/chemokine content by multiplex technology and relative changes are given in the graphs as percentage compared to no-RNP controls. **C)** Frequency of neutrophils migrated towards the supernatant of keratinocytes stimulated with recombinant IL-17A after *CEBPB* or *PTTG2* knock-out compared to IL-17A-stimulated wild type keratinocytes. **D)** Knockout of *CEBPB* or *PTTG2 in* neutrophil-like cells (differentiated HL60 cells) and consecutive migration towards CXCL8 or supernatant of wild type keratinocytes (wt KC SN) stimulated with recombinant IL-17A. KC: keratinocyte; RNP: ribonucleoprotein; SN: supernatant.

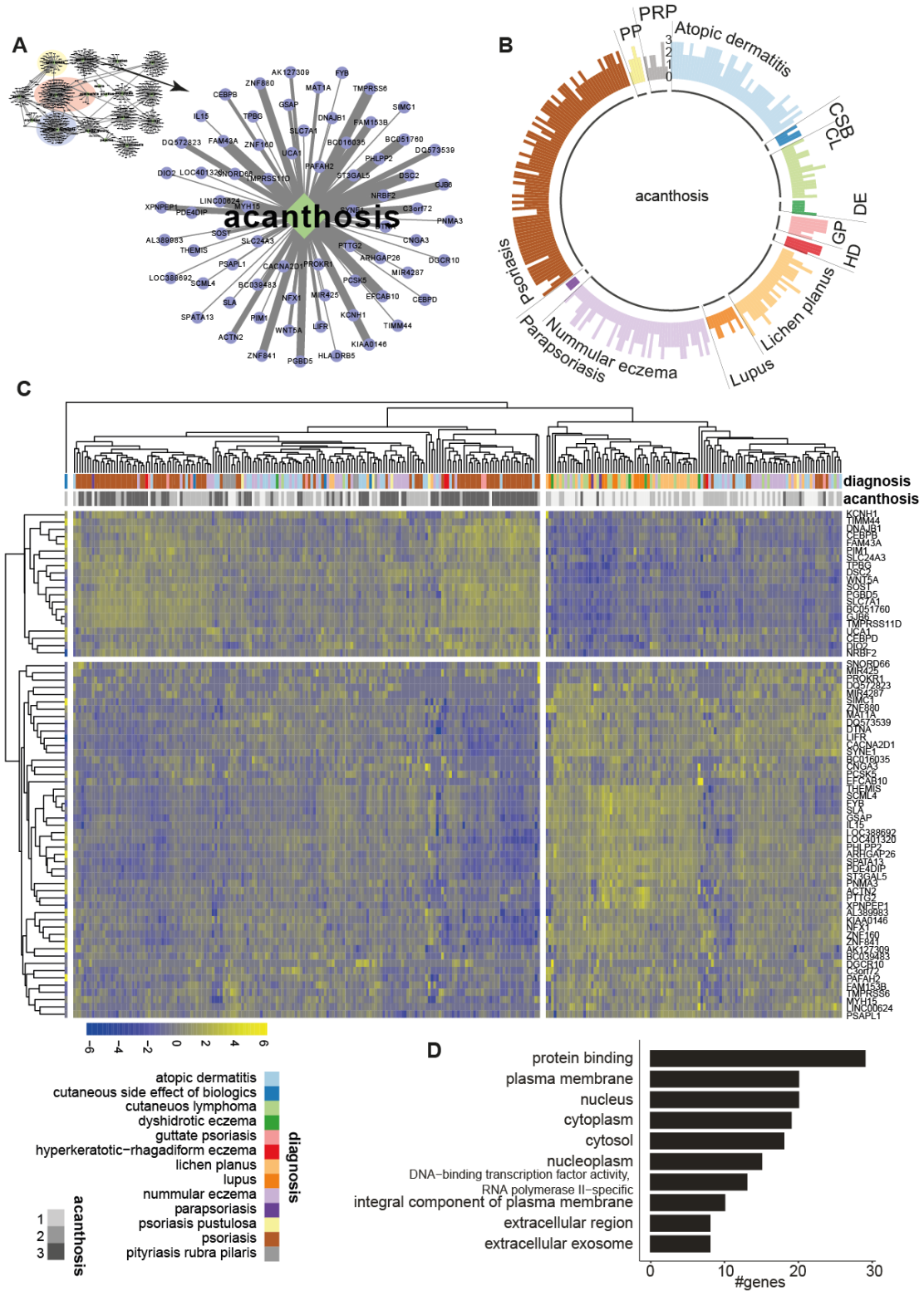

Fig. 5. Gene expression landscape of acanthosis. **A)** A focus on the clinical attribute 'acanthosis' and its associated genes. Line thickness indicates the strength of the gene-attribute association determined by the presence of the association in five iterating bio-computational cycles (thickest line=five times present). **B)** Distribution of the attribute 'acanthosis' across the 13 investigated non-communicable inflammatory skin diseases. Each bar represents one patient, and the height of the bar indicates the level of acanthosis measured by skin histology, with 0 = no, 1 = low-, 2 = moderate, 3 = strong acanthosis (see also Fig. S2). **C)** Hierarchical clustering of acanthosis-associated genes assigned with diagnosis (colors) and strength of acanthosis (grayscale). **D)** Gene ontology terms associated with the acanthosis gene set. CL: cutaneous lymphoma; CSB: cutaneous side effect of biologics; DE: dishydrotic eczema; GP: guttate psoriasis; Parapso: parapsoriasis; PP: psoriasis pustulosa; PRP: pityriasis rubra pilaris.

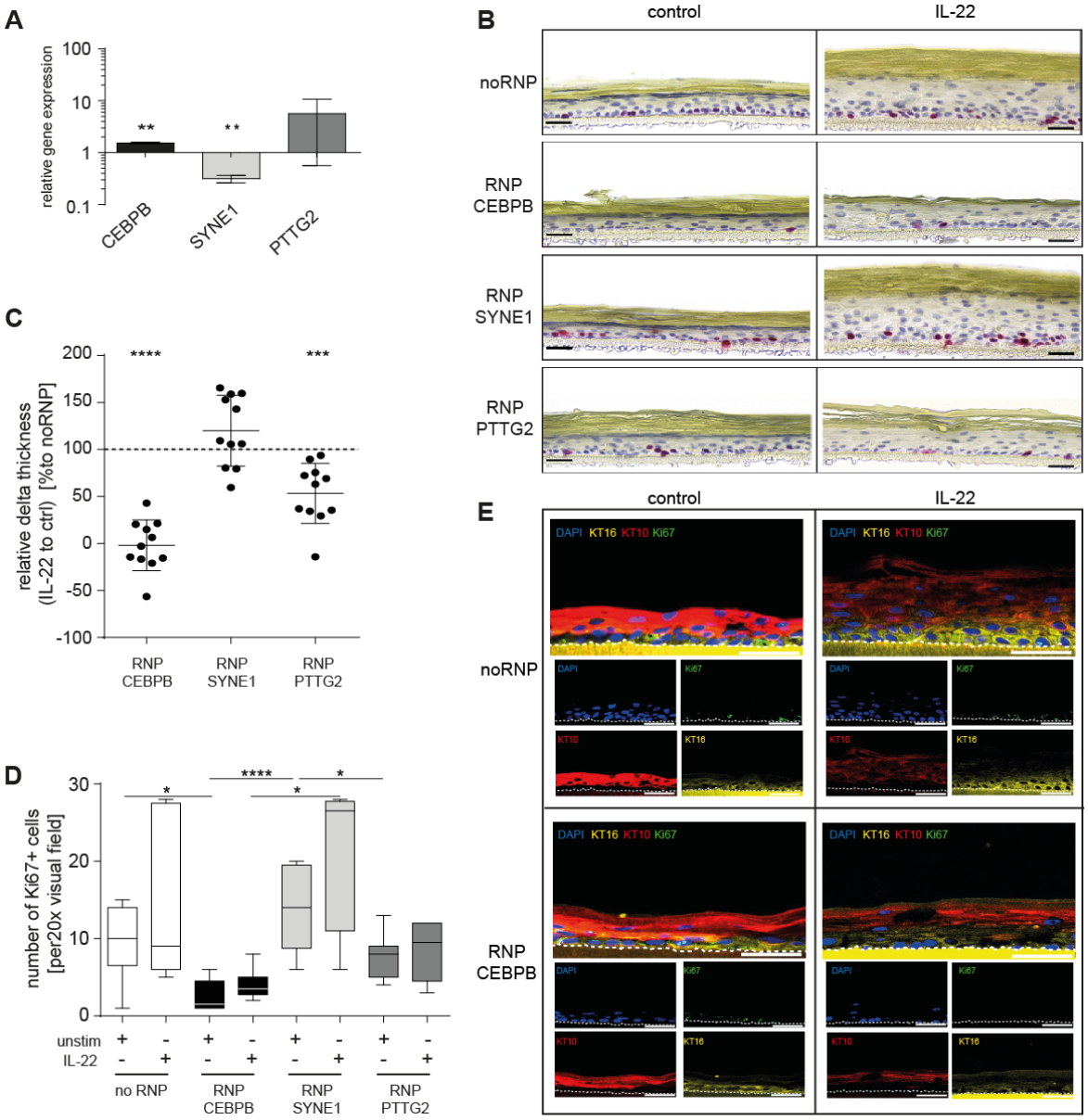

Fig. 6. CEBPB, SYNE1, and PTTG2 are mutually antagonistic regulators of acanthosis. **A)** Genes with a strong and yet unknown association to the attribute acanthosis were chosen and their expression was evaluated in primary human keratinocytes after IL-17A and TNF-α stimulation by real time PCR. **B)** The acanthosis-associated genes *CEBPB, SYNE1 and PTTG2* were knocked out in primary human keratinocytes using CRISPR-Cas9. 3D skin models of knock-out (RNP) and wild-type (no-RNP) were stimulated with recombinant IL-22 to induce acanthosis or left untreated (control) and stained for the proliferation marker Ki67. Representative stainings are shown. Bars represent 40µm. **C)** Relative delta thickness of 3D skin models with RNP-CRISPR knock-out for *CEBPB*, SYNE1 or *PTTG2* (n = 11). Here, thickness of each IL-22 stimulated model was normalized to its untreated control followed by calculation of percental changes of RNP vs no-RNP treated models. **D)** Quantification of Ki67 positive cells in 3D skin models shown in B). **E)** Representative Keratin 10 (KT10) (red), Keratin 16 (KT16) (yellow), and Ki67 (green) immunofluorescence staining of 3D skin models with a RNP-CRISPR knock out of *CEBPB* with or without (control) IL-22 stimulation. Nuclei were stained with DAPI (blue). The dotted white line indicates the border between basal keratinocytes and the culture membrane. Results on SYNE1 and PTTG2 are given in Fig. S8. Bars represent 50 µm. KT: kerat