Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways

Michela Palmieri1, Soren Impey2, Hyojin Kang1, Alberto di Ronza1, Carl Pelz2, Marco Sardiello1,∗ and Andrea Ballabio1,3,4,∗

1Department of Molecular and Human Genetics, Baylor College of Medicine, Jan and Dan Duncan Neurological Research Institute at Texas Children’s Hospital, Houston, TX, USA, 2Oregon Stem Cell Center, Oregon Health and Sciences University, Portland, OR, USA, 3Telethon Institute of Genetics and Medicine (TIGEM), Naples, Italy and 4Department of Pediatrics, Federico II University, Naples, Italy

Received May 18, 2011; Revised July 8, 2011; Accepted July 8, 2011

In metazoans, lysosomes are the center for the degradation of macromolecules and play a key role in a variety of cellular processes, such as autophagy, exocytosis and membrane repair. Defects of lysosomal pathways are associated with lysosomal storage disorders and with several late onset neurodegenerative diseases. We recently discovered the CLEAR (Coordinated Lysosomal Expression and Regulation) gene network and its master gene transcription factor EB (TFEB), which regulates lysosomal biogenesis and function. Here, we used a combination of genomic approaches, including ChIP-seq (sequencing of chromatin immunoprecipitate) analysis, profiling of TFEB-mediated transcriptional induction, genome-wide mapping of TFEB target sites and recursive expression meta-analysis of TFEB targets, to identify 471 TFEB direct targets that represent essential components of the CLEAR network. This analysis revealed a comprehensive system regulating the expression, import and activity of lysosomal enzymes that control the degradation of proteins, glycosaminoglycans, sphingolipids and glycogen. Interestingly, the CLEAR network appears to be involved in the regulation of additional lysosome-associated processes, including autophagy, exo- and endocytosis, phagocytosis and immune response. Furthermore, non-lysosomal enzymes involved in the degradation of essential proteins such as hemoglobin and chitin are also part of the CLEAR network. Finally, we identified nine novel lysosomal proteins by using the CLEAR network as a tool for prioritizing candidates. This study provides potential therapeutic targets to modulate cellular clearance in a variety of disease conditions.

INTRODUCTION

Cell metabolism is controlled by complex networks of genes, proteins and metabolites, which sense the cellular environment and organize the appropriate responses. In the past two decades, the importance of regulatory gene networks in cell metabolism has become evident in every aspect of cell function. In this context, the transcriptional networks associated with the biogenesis and function of organelles are of particular interest. Examples include very general processes such as the regulation of mitochondrial energy metabolism (1,2), peroxisomal remodeling and biogenesis (3), endoplasmic reticulum and Golgi stress responses (4–7) and more specialized processes such as maturation of melanosomes (8), exocrine granule maturation (9), apical transport during enterocyte development (10) and secretion in neurons (11). These regulatory systems are typically based on master transcription factors (TFs) that bind to regulatory signals disseminated throughout genes with key roles in these processes. To unravel the target gene network of a given TF, it is therefore essential to identify its interaction sites (cis-regulatory motifs) and downstream target genes.

A variety of essential cellular pathways are associated with lysosomal function, including clearance of substrates, autophagy, exocytosis and cell membrane repair. Lysosomal dysfunction is associated with lysosomal storage disorders.

∗To whom correspondence should be addressed at: Telethon Institute of Genetics and Medicine (TIGEM), Via Pietro Castellino 111, 80131 Naples, Italy. Tel: +39 0816132207; Fax: +39 0815790919; Email: ballabio@tigem.it (A.B.); Department of Molecular and Human Genetics, Baylor College of Medicine, Jan and Dan Duncan Neurological Research Institute, 1250 Moursund St., Suite N1325, Houston, TX 77030, USA. Tel: +1 8328248871; Fax: +1 7137988728; Email: sardiell@bcm.edu (M.S.)

© The Author 2011. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
(LSDs), a group of more than 50 recessive inherited diseases in which the deficiency of a single protein participating in lysosomal function results in the disruption of specific degradative pathways followed by the progressive accumulation of undegraded substrates. LSDs are individually rare, but their prevalence as a group is significant, with an incidence of \( \approx 1:5000 \) live births (12). In addition to LSDs, there is increasing evidence that lysosomes play a role in the pathogenesis of common neurodegenerative diseases such as Alzheimer’s (13), Parkinson’s (14) and Huntington’s (15), due to their involvement in autophagic clearance pathways. These studies underline the importance of the lysosome as a central player in cell metabolism. Hence, the characterization of genes participating in lysosomal biogenesis and function is a critical step toward the understanding of basic processes in cell biology and pathogenic mechanisms in human disease (16).

We recently identified transcription factor EB (TFEB), a member of the basic helix-loop-helix leucine-zipper family of TFs, as a master regulator of lysosomal function (17). We showed that TFEB promotes the transcription of several lysosomal genes by direct binding to specific E-box sites at their promoters. We named this gene network CLEAR (Coordinated Lysosomal Expression and Regulation). Cells overexpressing TFEB have a larger number of lysosomes and an enhanced degradative capability against lysosomal substrates such as glycosaminoglycans and autophagy substrates such as protein aggregates (17). In a more recent study, we also showed that TFEB directly regulates the biogenesis of autophagosomes, the fusion of autophagosomes with lysosomes and the autophagic flux (18). Such integrated control of the biogenesis of two cellular organelles that are involved in a tight partnership to promote cellular catabolism implies the coordination of a complex array of molecules and processes, from the synthesis of lysosomal and autophagosomal constituents and their correct localization into the cell to their physical or functional interaction (19). Therefore, the characterization of the TFEB targetome is of crucial importance to gain insight into the genetic control of these processes and their mutual relationships.

We previously performed transient TFEB overexpression in HeLa cells and profiled gene expression by using microarray technologies (17). Only a small portion of the genes whose expression was significantly induced in this system encoded lysosomal proteins, suggesting that TFEB activated a variety of different cellular processes. Here, we combined microarray data, deep sequencing of TFEB chromatin immunoprecipitate (ChIP-seq) and unbiased genomic and expression meta-analysis to obtain a more comprehensive map of the TFEB targetome. The results revealed a control system by which TFEB coordinates the expression of genes involved in the early and late steps of lysosomal biogenesis, its coordination with endosomal, trans-Golgi and autophagic pathways and the connection with non-lysosomal degradative pathways of the cell. We also utilized the CLEAR network to pinpoint candidate lysosome-resident proteins that underwent experimental validation, leading to the identification of nine novel lysosomal proteins.

**RESULTS**

**TFEB ChIP-seq analysis**

We performed ChIP using HeLa stable transfectants expressing TFEB-3xFLAG. Two independent ChIP samples were tested by real-time quantitative reverse transcription polymerase chain reaction (qRT–PCR), which showed a strong enrichment (>100-fold) for lysosomal promoters containing CLEAR sites compared with housekeeping genes and to immunoglobulin G (IgG) control samples obtained from
HeLa cells not expressing TFEB-3xFLAG (Fig. 1A). Two ChIP samples and two IgG control samples were subjected to Illumina high-throughput sequencing and mapped to the reference human genome. The number of sequenced and mapped tags analyzed is shown in Supplementary Material, Table S1. As the TFEB ChIP-seq replicates had a high degree of overlap (63%; 1129/1781), the two experiments were combined. A sliding window algorithm identified 6140 ChIP-seq peaks at a false discovery rate of 0.01. Importantly, TFEB ChIP-seq peaks accumulated at the 5' end of annotated RefSeq genes, whereas IgG control data showed no such enrichment (Supplementary Material, Fig. S1). One thousand six hundred and sixty-one peaks (27%) mapped within 10 kb from the transcription start site (TSS) of the closest gene, with a strong concentration at <1 kb from the TSS (821, 13%; Fig. 1B). To gain insight into the functional significance of the distribution of TFEB peaks, we compared ChIP-seq data with expression changes associated with TFEB transient overexpression in HeLa cells (17) and observed a significant overlap between up-regulated genes and TFEB peaks located within 1 kb from TSS ($P < 10^{-4}$), whereas this overlap was not significant for more distal TFEB peaks (Fig. 1C). Moreover, gene ontology (GO) analysis showed that the first group of genes was highly enriched for lysosomal genes (15 of 59, Bonferroni-adjusted $P$-value $10^{-11}$), whereas up-regulated genes with distal TFEB peaks did not show any significantly enriched category. These data strongly suggest that only the subset of TFEB binding sites associated with proximal promoters is functional. However, it cannot be excluded that a subset of distal sites can be functional in vivo in particular tissues or conditions.

Identification of TFEB direct targets

Next, we performed a de novo motif analysis of all high-confidence peaks using the Weeder webtool (20,21). This analysis revealed a strong enrichment for the consensus sequence TCACGTGA (Fig. 2A). We next examined the probability that this motif occurs within a 500-bp window including TFEB ChIP-seq peaks ranked by tag content. More than 80% of the top 200 ChIP-seq peaks contained a motif with a significance of approximately 20% over the next 800 ranked peaks (Supplementary Material, Fig. S2). A partial CLEAR motif (TCACG) showed a significant enrichment over a wide range of peaks, suggesting the existence of two classes of TFEB sites (Supplementary Material, Fig. S2). Analysis of CLEAR motifs in H1P promoters confirmed their tight association with TSS, with most sites lying in positions from −300 to +100 bp from the 5' end of genes (Fig. 2B). In order to expand the list of putative TFEB direct targets, we performed a genomic scan of CLEAR sites by using a position weight matrix (PWM) for TFEB target sequences based on the CLEAR motifs found in TFEB ChIP peaks. We focused our analysis on the regions spanning from −300 to +100 bp from all annotated TSSs of human genes. We found that 3972 promoters, corresponding to 3468 genes, contained CLEAR-like sequences in this region. In 7% of the cases, two or more CLEAR sites were detectable. Multiple CLEAR sites were found in 35% of genes positive to ChIP-seq and up-regulated by TFEB overexpression, indicating that multiple sites are best associated with the TFEB-mediated regulation of these genes. The analysis of the relative position of CLEAR sites showed that multiple sites tend to be tightly clustered (Fig. 2C). Several different types of promoter configurations can be described based on CLEAR site relative positioning, including single or randomly spaced multiple sites and CLEAR tandems or tetramers placed upstream of the TSS. CLEAR tandems and tetramers show the highest occurrence in genes known to be involved in lysosomal metabolism (Fig. 2D), although obvious subfunctional categories could not be detected.
Genes sharing a common upstream transcriptional control should display similar expression behavior in multiple conditions. Therefore, we examined the expression of TFEB direct targets in a variety of experiments to infer additional genes showing a similar expression profile. Genes with similar behavior and CLEAR sites in their promoters are likely to represent bona fide TFEB direct targets (Fig. 3A).

To investigate the transcriptional behavior of TFEB candidate target genes, we analyzed a vast set of published microarray experiments available at the Gene Expression Omnibus (GEO) database (22). Multiple cellular conditions and tissues are represented in this database. To ensure data homogeneity, we focused on Affymetrix platform HG-U133 Plus 2.0, for which 105 different experimental data sets were available at the time of the analysis. We first focused on the 59 genes induced by TFEB overexpression with TFEB ChIP-seq peaks adjacent to their transcriptional start sites. For each gene pair, we calculated a pair-wise co-expression score equal to their cumulative occurrence in the top 3% of correlated genes across all investigated experiments (see Material and Methods for details). Notably, pair-wise scores of these 59 TFEB targets were on average much higher than those of random sets of genes (Fig. 3B), thus indicating that such a correlation system was suitable to detect co-expression among genes of the TFEB/CLEAR network.

To investigate in more detail the expression relationships among these ‘core’ TFEB targets, we performed hierarchical clustering based on their co-expression scores. The results

Figure 3. Identification of TFEB direct targets. (A) Schematic of the rationale of the approach for integrating TFEB ChIP-seq and overexpression data with promoter sequence analysis and co-expression meta-analysis. (B) Frequency distribution of the scores associated with the co-expression analysis of putative TFEB targets and random genes. (C) Heat map of the scores associated with the co-expression analysis of putative TFEB targets. A cluster of genes with strongly associated expressions is indicated (red box). (D) Venn diagram showing overlapping genes in the analyzed data sets. Numbers in red indicate the over-representation of lysosomal genes as fold change. (E) Over-representation of lysosomal genes along the first 1000 genes obtained by the rank product of the four data sets (see text). The red line indicates a fold-change of 3.
showed that the vast majority of these genes (42 of 59, 71%) are grouped in a single cluster of strongly co-expressed genes, whereas the remaining genes showed weak or no expression correlation (Fig. 3C). Interestingly, the first group was highly enriched with lysosomal genes, whereas the latter did not contain any known gene associated with lysosomal function. The genes that did not show expression correlation could be false positives of the ChIP-seq analysis or, alternatively, they could belong to other gene networks that require additional TFs in combination with TFEB for the proper modulation of their expression.

To infer additional genes belonging to the CLEAR network, we looked for genes co-expressed with the main cluster (boxed in red in Fig. 3C) by extending the analysis to all genes represented in the HG-U133. We flagged genes with an average score of at least 8 against the members of this cluster, a cutoff score that eliminates 95% of interactions among randomly chosen genes (Fig. 3B). Of this group of co-expressed genes, 545 had CLEAR sites in their core promoters and are therefore prime candidates for being TFEB targets. The comparison of the lists of genes inferred by ChIP-seq, TFEB overexpression and sequence/co-expression analysis showed that 36 genes were shared among all lists. This gene list represents our most conservative estimate of TFEB direct targets and includes 19 genes already known to be involved in lysosomal/autophagic pathways, with a considerable enrichment compared with random lists of the same size (Bonferroni-adjusted P-value $\leq 10^{-13}$; Fig. 3D). Additional 145 genes were inferred by any combination of two of the methods; these genes also are significantly enriched for lysosomal functions, albeit in different proportions according to the combined lists (Fig. 3D).

These results suggest that each list alone is far from completion and that their combination overall improves the identification of putative TFEB targets. However, it is possible that the thresholds imposed to the various lists contribute to the exclusion of true targets. Thus, we decided to use the rank product method to combine the full lists of genes obtained by ChIP-seq, TFEB overexpression, promoter analysis and co-expression meta-analysis. In the rank product method, all gene lists contribute to the final sorting of genes by means of their respective gene rankings. The genes that rank high in multiple lists are more likely to rank high in the final rank product list, even if they rank low in some of the lists (23). The advantage of the rank product method is that there is no need for a priori thresholds to select members of the single lists: a cutoff can be applied in the resulting combined list by examining variations in composition and properties along its ranked genes.

We first ranked the four gene lists according to: (i) peak height in TFEB ChIP-seq, (ii) fold-induction in TFEB overexpression, (iii) score to TFEB core target genes in co-expression meta-analysis and (iv) number of CLEAR sites and their CLEAR PWM scores in promoter analysis (see Materials and Methods for details). Rank product of the four lists resulted in a new list, which was further enriched for lysosomal genes in top-ranking positions than any of the four previous lists alone (data not shown). Given the specificity of TFEB for lysosomal pathways, we used the over-representation of lysosomal genes as a criterion to survey the combined list. This analysis showed a significant enrichment for lysosomal genes up to position 500 (Fig. 3E). Therefore, we set this as a threshold for the final list of TFEB putative functional targets and removed genes inferred by only one of the four analyses ($N = 29$). Out of the remaining 471 genes, 435 (92%) were found to have CLEAR sites within $-300$ to $+100$ bp from their TSS; 368 (78%) had TFEB peaks at $<1$ kb from TSS in ChIP-seq analysis; 226 (48%) were found to be co-expressed with ‘core’ TFEB targets in co-expression meta-analysis and 106 (23%) were induced by TFEB overexpression in HeLa cells.

### Biological pathways controlled by the CLEAR network

Genes with a known role in lysosomal biogenesis and function that are included in the rank product list are summarized in Table 1 and represent the most likely lysosomal direct targets of TFEB. These genes encode for proteins that can be grouped in several distinct categories, including lysosomal hydrolases and accessory proteins, lysosomal membrane proteins, subunits of the proton pump, proteins participating in autophagy and non-lysosomal proteins involved in lysosomal biogenesis (Fig. 4A).

To identify the relevant pathways controlled by TFEB, we performed GO analysis of the 471 TFEB targets looking for over-represented classes of genes. The analysis of ‘cellular compartment’ terms showed that enrichments for classes linked to the lysosome were the most significant ones. However, also components associated with the endosome, mitochondrion and Golgi complex were significantly over-represented (Supplementary Material, Table S2). Albeit with no over-representation, virtually all other organelles and cell parts contributed to the set of genes controlled by TFEB (Fig. 4B).

The analysis of ‘biological function’ terms showed that genes involved in lysosomal catabolism were the most enriched among TFEB targets. However, genes involved in endocytosis, protein transport, cellular response to stress and carbohydrate metabolism were also over-represented in this gene set (Supplementary Material, Table S2). A considerable proportion of genes did not

### Table 1. TFEB direct targets with a known role in lysosomal function

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysosomal hydrolases and accessory proteins</td>
<td>ASAHI1, CTS4, CTSB, CTSID, CTSF, GAA, GALNS, GBA, GLA, GLBI, GNS, GUSB, HEXA, HFXB, IFIT3, NAGLU, NEU1, PLBD2, PPT1, PSAP, SCPEP1, SGSH, TPP1</td>
</tr>
<tr>
<td>Lysosomal membrane</td>
<td>C1orf85, CD63, CLCN7, CLN3, CTNS, MCOLN1, SLC36A1, LAMP1, TME55B</td>
</tr>
<tr>
<td>Lysosomal acidification</td>
<td>ATP6AP1, ATP6V0A1, ATP6V0B, ATP6V0C, ATP6V0D1, ATP6V0D2, ATP6V0E1, ATP6V1A, ATP6V1B2, ATP6V1C1, ATP6V1D, ATP6V1E1, ATP6V1G1, ATP6V1H</td>
</tr>
<tr>
<td>Non-lysosomal proteins involved in lysosomal biogenesis</td>
<td>NAGPA, GNPTG, IGF2R, M6PR, BLOC1S1, BLOC1S3, HPS1, HPS3, HPS5, SUMF1</td>
</tr>
<tr>
<td>Autophagy</td>
<td>BECN1, GABARAP, HIF1A, NRRF2, PRKAG2, RAB7A, RARGC, SQSTM1, STK4, UVRAG, VPS8, VPS11, VPS18, VPS26A, VPS33A, VPS35, WDR45</td>
</tr>
</tbody>
</table>
belong to any specific functional category, and about two-thirds of genes were associated with a variety of functions that are not obviously connected to either the lysosomal/autophagic function or the other over-represented categories on the basis of the data available to date (Fig. 4C).

To map in detail the TFEB-controlled cellular pathways, we analyzed the list of TFEB targets by using the ‘Kyoto Encyclopedia of Genes and Genomes’ (KEGG) database (24). The analysis showed that a number of biochemical and cellular pathways are targeted by TFEB, which cover a significant proportion of cell metabolism (Table 2).

**Identification of novel lysosomal proteins**

To date, only ~100 lysosomal proteins have been identified and more than half of them are associated with diseases in human or animal models. In recent years, several algorithms have been developed for the prediction of the subcellular localization of eukaryotic proteins (25,26). Among these, SherLoc provides parameters relative to the lysosome (27). In particular, it assigns a support vector machines (SVMs) score to input proteins, which reflects their overall amino acid composition compared with the composition of proteins known to reside in the various cellular organelles. SVMaac scores range from 0.5 to 1 for amino acid compositions compatible with the selected organelle and from −1 to −0.5 for putative incompatibility.

We confirmed that the SVMaac score is reliable for lysosomal proteins by analyzing the full collection of human proteins: we found that 80% of established lysosomal proteins have a lysosomal SVMaac score above 0.5, compared with the 12% of the whole human proteome (Fig. 5A). Moreover,
The list of human proteins ranked according to their lysosomal SVMaac scores is highly enriched for known lysosomal proteins at top positions, with 50% of established lysosomal proteins lying within the first ~300 positions (Fig. 5B).

To indentify novel candidate lysosomal proteins within the list of putative TFEB targets, we calculated the rank product between the two lists of human genes, ranked according to lysosomal SVMaac scores of their encoded proteins and TFEB/CLEAR network scores, respectively. This resulted in a further enrichment of lysosomal genes at top positions of the product rank list (Fig. 5C). Therefore, top positions are also likely to contain genes encoding novel lysosomal proteins (see below).

To increase the number of candidates, we also considered genes not belonging to the CLEAR network but still with an expression correlated with other lysosomal genes. To this

<table>
<thead>
<tr>
<th>KEGG term</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biochemical pathways</td>
<td></td>
</tr>
<tr>
<td>hsa00010: glycolysis/gluconeogenesis</td>
<td>ALDOA, AKR1A1</td>
</tr>
<tr>
<td>hsa00020: citrate cycle (TCA cycle)</td>
<td>SDHC, OGDH, OGDH</td>
</tr>
<tr>
<td>hsa00051: fructose and mannose metabolism</td>
<td>ALDOA, PKFB2</td>
</tr>
<tr>
<td>hsa00062: fatty acid elongation in mitochondria</td>
<td>PPT1</td>
</tr>
<tr>
<td>hsa00100: steroid biosynthesis</td>
<td>LSS, SC5DL</td>
</tr>
<tr>
<td>hsa00190: oxidative phosphorylation</td>
<td>UQCC2, ATP5D, ATP5J2, SDHC, ATP5I</td>
</tr>
<tr>
<td>hsa00240: pyrimidine metabolism</td>
<td>UPP1, ENTPD6, TNNRD1, POLR3C</td>
</tr>
<tr>
<td>hsa00270: cysteine and methionine metabolism</td>
<td>AHCTYL2</td>
</tr>
<tr>
<td>hsa00310: lysine degradation</td>
<td>PLOD3, OGDHL, SUV3H1, OGDH</td>
</tr>
<tr>
<td>hsa00480: glutathione metabolism</td>
<td>GSR, GPX1, GSTO1</td>
</tr>
<tr>
<td>hsa00500: starch and sucrose metabolism</td>
<td>GBE1, GUSB, GAA, UGDH</td>
</tr>
<tr>
<td>hsa00510: N-glucan biosynthesis</td>
<td>MGA4B, ALG1, RPNI</td>
</tr>
<tr>
<td>hsa00511: other glycan degradation</td>
<td>HEXA, HEXB, EU1, FUC2, GLB1, GBA</td>
</tr>
<tr>
<td>hsa00512: O-glucan biosynthesis</td>
<td>ST3GAL1</td>
</tr>
<tr>
<td>hsa00520: amino sugar and nucleotide sugar metabolism</td>
<td>AMDHD2, GNPD51, HEXA, HEXB, UGDH, NAGK</td>
</tr>
<tr>
<td>hsa00531: glycosaminoglycan degradation</td>
<td>SGSH, GNS, NAGLU, HEXA, GUSB, HEXB, GALNS, GLB1</td>
</tr>
<tr>
<td>hsa00600: sphingolipid metabolism</td>
<td>GLA, NEU1, ASAH1, GLB1, GBA</td>
</tr>
<tr>
<td>hsa00604: glycosphingolipid biosynthesis</td>
<td>ST3GAL1, SLC33A1, HEXA, HEXB, GLB1</td>
</tr>
<tr>
<td>hsa00670: one carbon pool by folate</td>
<td>SHMT1, MTHFR</td>
</tr>
<tr>
<td>hsa00770: pantothenate and CoA biosynthesis</td>
<td>PPRC5</td>
</tr>
<tr>
<td>hsa00860: porphyrin and chlorophyll metabolism</td>
<td>HMOX1, GUSB, BVRB, UROD</td>
</tr>
<tr>
<td>hsa00900: terpeneind backbon biosynthesis</td>
<td>DHD1S</td>
</tr>
<tr>
<td>Cellular processes associated with lysosomal function</td>
<td></td>
</tr>
<tr>
<td>hsa04130: SNARE interactions in vesicular transport</td>
<td>STX4, STX6</td>
</tr>
<tr>
<td>hsa04140: regulation of autophagy</td>
<td>BECN1, GABARAP</td>
</tr>
<tr>
<td>hsa04144: endocytosis</td>
<td>ARFGAP1, CLTA, RAB5A, GIT2, PDCD6IP, CLT, CSF1R</td>
</tr>
<tr>
<td>hsa04150: mTOR signaling pathway</td>
<td>MAPK1, HIF1A, EIF4E, MAPK3</td>
</tr>
<tr>
<td>hsa04466: Mcl gamma R-mediated phagocytosis</td>
<td>MAPK1, MAPK3, CFL1, ARPC5</td>
</tr>
<tr>
<td>hsa04612: antigen processing and presentation</td>
<td>HLA-B, CTSB, CALR, GILT</td>
</tr>
<tr>
<td>hsa04916: melanogenensis</td>
<td>DVR2, MAPK1, MAPK3</td>
</tr>
<tr>
<td>Signaling pathways</td>
<td></td>
</tr>
<tr>
<td>hsa04010: MAPK signaling pathway</td>
<td>MAX, MAPK1, DUSP3, ATF4, DUSP1, PDGFB, RELB, MAPK3, DUSP10, MAPKAP3, STK4 MAPK1, GNB2, MAPK3</td>
</tr>
<tr>
<td>hsa04062: chemokine signaling pathway</td>
<td>BAX, CDK4</td>
</tr>
<tr>
<td>hsa04115: p53 signaling pathway</td>
<td>MAPK1, TBK1, LY96, MAPK3</td>
</tr>
<tr>
<td>hsa04262: TLR signaling pathway</td>
<td>MAPK1, MAPK3, TNFAP1, TRIP6</td>
</tr>
<tr>
<td>hsa04622: RLR signaling pathway</td>
<td>TBK1, TRIM25</td>
</tr>
<tr>
<td>hsa04630: Jak-STAT signaling pathway</td>
<td>PIA54, SOCS5</td>
</tr>
<tr>
<td>hsa04722: neurotrophin signaling pathway</td>
<td>MAPK1, ATF4, BAX, MAPK3, NGFRAP1, ARHGHDIA</td>
</tr>
<tr>
<td>hsa04910: insulin signaling pathway</td>
<td>MAPK1, EIF4E, PHKG2, PRKAG2, MAPK3, PPARGCA</td>
</tr>
<tr>
<td>hsa04920: adipocyte kinase signaling pathway</td>
<td>PRKAG2, PPARGC1A, CAMK1</td>
</tr>
<tr>
<td>DNA metabolism</td>
<td></td>
</tr>
<tr>
<td>hsa03030: DNA replication</td>
<td>MCM2</td>
</tr>
<tr>
<td>hsa03410: base excision repair</td>
<td>TDG, NTTL1, APEX1</td>
</tr>
<tr>
<td>hsa03420: nucleotide excision repair</td>
<td>RAD23A, CK7, GTF2H1</td>
</tr>
<tr>
<td>hsa03440: homologous recombination</td>
<td>TOP3B</td>
</tr>
<tr>
<td>hsa04623: cytosolic DNA-sensing pathway</td>
<td>TBK1, POLR3C, ADAR</td>
</tr>
<tr>
<td>Cell metabolism</td>
<td></td>
</tr>
<tr>
<td>hsa03010: ribosome</td>
<td>HNRNP2H2, RPS28, UBA52</td>
</tr>
<tr>
<td>hsa03022: basal TFs</td>
<td>GTF2A1, TAF6L, GTF2H1</td>
</tr>
<tr>
<td>hsa03040: spliceosome</td>
<td>HNRNP3, SNRP, SMNDC1</td>
</tr>
<tr>
<td>hsa04810: regulation of actin cytoskeleton</td>
<td>GNA13, MAPK1, PDGFB, MAPK3, CFL1, ARPC5</td>
</tr>
<tr>
<td>hsa04512: ECM–receptor interaction</td>
<td>DAG1</td>
</tr>
<tr>
<td>Other processes</td>
<td></td>
</tr>
<tr>
<td>hsa04710: circadian rhythm</td>
<td>BHLHE40, PER3</td>
</tr>
</tbody>
</table>
aim, we first defined a set of 98 well-characterized genes encoding proteins with exclusive or main lysosomal localization (Supplementary Material, Table S3). Then, to minimize redundancy with the TFEB/CLEAR list, we excluded from this set genes that are part of the CLEAR network according to our study. Subsequently, we performed co-expression meta-analysis of the remaining lysosomal genes to assign pairwise scores against any other genes represented in the HG-U133 platform, similarly to what we performed for CLEAR genes. We then used these average pair-wise scores to rank all genes included in the HG-U133 platform. Despite the exclusion of lysosomal CLEAR genes from the starting pool, this list resulted to be highly enriched for lysosomal genes at top positions (Fig. 5D). We again combined this gene list with the list ranked according to lysosomal SVMaac scores by the rank product method, which resulted in a new list of genes with a greater enrichment of known lysosomal genes at top-ranking positions (Fig. 5D).

From the two lists, we selected 20 high-ranking genes based on the absence of definitive data regarding the subcellular localization of their encoded proteins, their broad expression in a variety of tissues and cell types as verified at Unigene website (http://www.ncbi.nlm.nih.gov/unigene) and, in some cases, their possible involvement in catabolic processes or vesicle trafficking as predicted by their domain composition (Supplementary Material, Table S4). To investigate the subcellular localization of selected candidates, we transfected their cDNAs in HeLa cells and performed confocal microscopy analysis using lysosomal marker LAMP2. The analysis showed that 9 of the 20 candidates (45%) co-localize with lysosomes in the tested conditions (Fig. 6). These are mostly transmembrane proteins with unknown function and have predicted size between 238 and 771 amino acids (Table 3). Among them, SLC26A11 has been recently shown to co-localize with lysosomal markers also by another study (28) based on TFEB overexpression data (17). These proteins will require a detailed biochemical analysis to establish their function and to understand their possible involvement in human disease.

DISCUSSION

Previous studies showed that TFEB is ubiquitously expressed (29) and that its deficiency causes early embryonic lethality in mouse (30), thus suggesting that TFEB has an essential, ubiquitous and non-redundant function. To define the TFEB targetome, we integrated ChIP-seq data with sequence analysis (i.e. search of TFEB cis-acting motifs genome-wide) and functional data (i.e. transcriptional response to TFEB overexpression and co-expression meta-analysis of lysosomal genes). It should be noted that a great variety of conditions underlay the acquisition of the experimental data: the TFEB ChIP-seq was performed in HeLa TFEB-3xFLAG stable clones in a steady-state condition; the TFEB overexpression was performed in HeLa cells in condition of transient transfection and the co-expression meta-analysis included dozens of different cell/tissue types and conditions (31). However, despite such different conditions, the results showed high consistency across the various data sets and a great overlap with the genomic cis-motif search, which is intrinsically independent on experimental conditions. This indicates that HeLa cells
are a suitable system for the experimental study of the relationship between TFEB and its targetome, even in non-physiological conditions (TFEB or TFEB-3xFLAG overexpression). Additional studies in different cell types, and in vivo in different tissues, will be needed to characterize the full spectrum of TFEB functions and the entire list of its targets.

Our analysis also showed that TFEB tends to bind to the core promoters of its targets genes, whereas no evidence is found for a functional binding of sequences more distant from the TSS. These data are in line with the increasing evidence that the regulators of cell metabolism promote gene transcription by binding DNA at the core promoter and likely directly interact with the basal transcription machinery (32). Our integrated study involved TFEB overexpression, ChIP-seq and cis-motif search analyses, co-expression meta-analysis and embraced a wide variety of metabolic conditions, thus may serve as a general model for the dissection of the targetomes of TFs involved in the regulation of the cellular metabolism. It would be of interest to discriminate among different transcript isoforms due to either alternative promoters or alternative splicing (33,34). To minimize the possible biases for specific isoforms due to probe specificity, in our co-expression meta-analysis, we used as ‘baits’ all available probes associated with all genes of interest (i.e. the putative TFEB targets inferred by ChIP-seq and TFEB overexpression). This way even if a single probe recognizes only a particular spliced form, the ensemble of all other probes (from the same gene and other genes) should guarantee an overall robust analysis, as shown by the high coherence among the sets of genes inferred by co-expression meta-analysis and those inferred by the other methodologies.

Several interesting features resulted from the analysis of the CLEAR network. TFEB targets control multiple sequential steps of at least five degradation pathways, including the degradation of heparan and keratan sulfate among glycosaminoglycans and of sphingolipids (Fig. 7). Surprisingly, no genes encoding enzymes for the degradation of dermatan sulfate, another type of glycosaminoglycan, appeared to be regulated by TFEB targets in our analysis. Although heparan sulfate is found virtually in all animal tissues and keratan sulfate is diffused in multiple tissues (including cartilage, bone, cornea and the central nervous system), dermatan sulfate is mainly found in skin, which likely demands a different control for its degradation. TFEB also targets genes encoding the subunits of the vacuolar proton pump, which creates and maintains the lysosomal acidic environment necessary to the lysosomal hydrolases for their hydrolytic reactions.

Another interesting feature is that TFEB regulates also extra-lysosomal degradative pathways: four of the five steps for the degradation of chitin are controlled by TFEB and occur in the cytosol (with the exception of the step catalyzed by hexosaminidase A, which is lysosomal). Degradation of hemoglobin, for which TFEB controls two sequential reactions, is also extra-lysosomal.

Moreover, among TFEB targets are genes encoding proteins associated with autophagy. Notable examples include BECN1 and UVRAG, which form a complex that promotes the formation of double-membraned organelles associated with autophagy proteins ATG5 and ATG12 (35,36); HIF1A, a TF shown to promote autophagy in chondrocytes (37,38) and in hypoxia conditions in cultured cells (39,40); RRAGC, member of a group of Rag GTPases that mediate amino acid signaling to mTORC1 (41) and the polyubiquitin-binding protein p62/SQSTM1, which facilitates autophagic degradation of ubiquitinated protein aggregates (42). Additional signaling pathways targeted by TFEB are the mitogen-activated protein (MAP) kinases and p53 signaling, both known to play key roles in cell proliferation, differentiation and survival/apoptosis (43). p53 has recently been associated with autophagy: in particular, it has been shown that p53 is activated by metabolic stress through the mammalian target of rapamycin (mTOR) signaling pathway, and in turn, it regulates autophagy to provide cell metabolites for surviving through a damage-regulated autophagy modulator (44). Interestingly, these data are consistent with our recent study demonstrating that TFEB directly regulates the biogenesis of

Figure 6. Experimental validation of novel lysosomal proteins. HeLa cells were transiently transfected with plasmids carrying cMyc-tagged versions of the indicated proteins and analyzed 48 h after transfection. After fixation, the lysosomal marker protein LAMP2 was visualized by indirect immunofluorescence using a specific monoclonal anti-LAMP2 antibody and a fluorescently labeled secondary antibody.
autophagosomes, the fusion between lysosomes and autophagosomes and the autophagic flux, thus representing a link between autophagy and lysosomal biogenesis (18).

Notably, TFEB-mediated regulation extends to non-lysosomal proteins involved in lysosomal biogenesis. Examples are GNPTG and NAGPA, which reside in the Golgi and catalyze the first and the second steps in the synthesis of the mannose 6-phosphate recognition marker on newly synthesized lysosomal hydrolases (45); and both cation-dependent and cation-independent mannose 6-phosphate receptors M6PR and IGF2R, which recognize and deliver the modified hydrolases to lysosomes. Also included in this group are BLOC1S1, BLOC1S3, HPS1, HPS3 and HPS5, which are components of the biogenesis of lysosome-related organelles complex-1, -2 and -3 (46–49) and SUMF1, which is essential for the activation of lysosomal sulfatases by the post-translational modification of a cysteine residue at their active site (50,51).

Among TFEB direct targets are genes belonging to distinct families of pattern recognition molecules including membrane-anchored Toll-like receptors (TLRs), cytosolic Nod-like receptors (NLRs) and Rig-I-like receptors (RLRs), which are involved in innate immune detection of danger signals and microbial motifs (52). This observation suggests that TFEB activation may be required to enhance the lysosomal system in response to pathogen invasions.

Other interesting TFEB targets are peroxisome proliferator-activated receptor-gamma coactivator-1α, which is involved in mitochondrial biogenesis and the insulin signaling pathway (1,53,54), PRKAG2, a member of the AMPK (MAP kinase) gamma subunit family involved in the regulation of the synthesis of cholesterol and fatty acids (55),

Table 3. Proteins co-localizing with LAMP2 in HeLa cells

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>RefSeq ID</th>
<th>Size (amino acid)</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPR137B</td>
<td>G protein-coupled receptor 137B</td>
<td>NP_003263</td>
<td>399</td>
<td>1q42.3</td>
</tr>
<tr>
<td>MARCH8</td>
<td>Membrane-associated ring finger (C3HC4) 8</td>
<td>NM_001002265</td>
<td>292</td>
<td>10q11.2</td>
</tr>
<tr>
<td>SERINC2</td>
<td>Serine incorporator 2</td>
<td>NP_849196</td>
<td>455</td>
<td>1p35.2</td>
</tr>
<tr>
<td>SLC37A3</td>
<td>Solute carrier family 37, member 3</td>
<td>NP_996996</td>
<td>494</td>
<td>7q34</td>
</tr>
<tr>
<td>TMEM8A</td>
<td>Transmembrane protein 8A</td>
<td>NP_067082</td>
<td>771</td>
<td>16p13.3</td>
</tr>
<tr>
<td>SLC26A11</td>
<td>Solute carrier family 26, member 11</td>
<td>NP_001159821</td>
<td>606</td>
<td>17q25.3</td>
</tr>
<tr>
<td>MFSD1</td>
<td>Major facilitator superfamily domain containing protein 1</td>
<td>NP_001161375</td>
<td>475</td>
<td>3q25.32</td>
</tr>
<tr>
<td>PLD3</td>
<td>Phospholipase D3</td>
<td>NP_036400</td>
<td>489</td>
<td>19q13.2</td>
</tr>
<tr>
<td>TMEM127</td>
<td>Transmembrane protein 127</td>
<td>NP_060391</td>
<td>238</td>
<td>2q11.2</td>
</tr>
</tbody>
</table>

Figure 7. KEGG reactions associated with TFEB direct targets. Reported diagrams show examples of degradation pathways in which multiple steps are catalyzed by enzymes included in the CLEAR network (shaded in orange). Degradation of glycosaminoglycans and sphingolipids occur in the lysosome, and the other reported pathways are extra-lysosomal.
and genes involved in DNA replication and repair and in the synthesis and function of the ribosome, spliceosome and actin cytoskeleton (Table 2). Overall, these data are suggestive of a complex interplay between master regulators of metabolism and a variety of cellular processes.

TFEB also appears to control the synthesis of phosphatidylinositol-4-phosphate (PI(4)P) and phosphatidylinositol-4,5-bisphosphate (PI(4,5)P2) (Supplementary Material, Fig. S3). These membrane phospholipids are enriched in the trans-Golgi network and in the plasma membrane, respectively. One of the target enzymes of this pathway, PI4KA, has been involved in the regulation of a pool of PI(4)P associated with store-operated Ca2+ entry (SORE), which in turn plays a key role in the exocytosis of lysosome-related organelles in cytotoxic lymphocytes (56,57). TEBF also targets the soluble NSF attachment protein receptor (SNARE) proteins, syntaxins 4 and 6. SNAREs are membrane components that are important for the fusion of transport vesicles with the plasma membrane (exocytosis) or with target compartments such as lysosomes and are involved in vesicular transport, endocytosis and phagocytosis (58). Among TFEB targets, STX4 encodes for a plasma membrane t-SNARE enriched in microdomains where exocytosis occurs. STX4 interacts with synaptotagmin VII to regulate lysosomal exocytosis, a ubiquitous process important for repairing of plasma membrane wounds (59); STX4 also positively influences granulocyte exocytosis (60). Recent work has demonstrated that STX4 defines an exocytic zone directing membrane fusion for post-synaptic plasticity, and its disruption impairs spine exocytosis at hippocampal synapses (58). Interestingly, another TFEB target, MCOLN1, has been previously associated with lysosomal exocytosis: MCOLN1 deficiency indeed impairs lysosomal exocytosis (61), whereas MCOLN1 gain of function can increase it (62). Therefore, TFEB may modulate lysosomal exocytosis by controlling the expression of STX4 and MCOLN1 and the pool of SORE-associated PI(4)P. Recent data obtained in our laboratory indicate that TFEB directly regulates lysosomal exocytosis (63). Another TFEB target, STX6, is a regulator of the protein trafficking machinery required for cell proliferation, adhesion and survival (64). Notably, we have recently demonstrated an impairment of lysosomal fusion and SNARE function in LSDs (65). An unexpected result from the analysis of the CLEAR network is that a number of biochemical pathways central to cell metabolism, such as the folate cycle, fructose and glutathione metabolism, glycan and cholesterol biosynthesis and oxidative drug targets are targeted by TFEB at multiple steps (Supplementary Material, Fig. S3 and Table 2).

Recently, considerable progress has been made in the systematic identification of lysosomal proteins by using proteomic approaches based on either the physical separation of lysosomal components (such as membranes) or the selective purification of proteins with specific features, such as the mannose-6-phosphate modification (66–74). In the present work, we performed genome-wide co-expression meta-analysis and amino acid composition analysis of all proteins encoded in the human genome to investigate their similarity with established lysosomal proteins. This led to the identification of nine novel lysosomal proteins out of 20 tested, a significant number when compared with the ~100 lysosomal proteins established to date. Interestingly, these new proteins are likely associated with lysosomal membranes, indicating that the complement of known lysosomal proteins residing into the lumen is at nearly completion and has been saturated by previous analyses, which were driven by the search of enzymes defective in lysosomal disorders and later complemented with proteomic studies (28). On the contrary, a relatively small number of lysosomal membrane proteins have been isolated to date, likely accounting for only a proportion of the total. Interestingly, among the newly identified proteins, MARCH8 belongs to a family of membrane-bound E3 ubiquitin ligases, enzymes that add ubiquitin to specific lysines of target membrane proteins to trigger their endocytic internalization and subsequent lysosomal degradation (75,76). Thus, MARCH8 could represent a new lysosomal function specific to the direct selection and targeting of membrane proteins that undergo lysosomal degradation. Also of interest, the mapping of the gene encoding new lysosomal protein TMEM8A to the same chromosomal region where the locus for microhydranecephaly (MHAC, MIM:605013) was assigned (77). In this disease, patients display the storage of mucopolysaccharides (78), a hallmark of several LSDs. Therefore, TMEM8A is an obvious candidate gene for mutation screening in patients with MHAC. Future work will be aimed at the dissection of the function of these newly identified lysosomal proteins and at the understanding of their possible involvement in human disease.

Finally, we and others have recently proposed that lysosomal enhancement could be used as a therapeutic strategy for the treatment of LSDs and neurodegenerative diseases (14,15,79,80). Key to this aim is the dissection of the cellular processes controlled by TFEB that may have relevance for the therapy of these diseases, including autophagy and vesicle trafficking in addition to specific degradative pathways. TFEB-mediated lysosomal enhancement may be used to slow, halt or reverse the accumulation of undegraded molecules in affected tissues, toward a therapeutic clearance of pathogenic aggregates or storage deposits that are found in these diseases. The dissection of the CLEAR network provides candidate genes and pathways that may mediate TFEB therapeutic effects. These genes may in turn represent novel therapeutic targets.

MATERIALS AND METHODS

Molecular biology

Full-length human cDNAs for candidate lysosomal proteins (Supplementary Material, Table S4) were PCR-amplified from cDNA libraries (Clontech) and cloned into the pcDNA3.1/c-Myc vector (Invitrogen). RNA samples were obtained using the RNeasy kit (Qiagen) according to the manufacturer’s instructions. RNA was quantified using the NanoDrop 8000 (Thermo Fisher). cDNA was synthesized using QuantiTect Reverse Transcription kit (Qiagen).

Cell culture, transfection and confocal analysis

HeLa cells were grown in Dulbecco’s modified Eagle’s medium (Euroclone), supplemented with 10% heat inactivated fetal bovine serum (Hyclone). Transfection of full-length cDNAs for candidate lysosomal proteins was performed by
using PolyFect Transfection Reagent (Qiagen) according to the manufacturer’s protocols. Transfected cells were grown on glass coverslips for 24 h, washed with phosphate buffered saline (PBS) containing 100 mM MgCl$_2$ and 100 mM CaCl$_2$ (PBS/Ca/Mg) and fixed with 4% paraformaldehyde (PFA; Sigma-Aldrich) for 10 min. After washing and quenching PFA with 50 mM NH$_4$Cl for 15 min, cells were washed with PBS and permeabilized in blocking buffer (0.05% saponin/0.2% BSA in PBS/Ca/Mg) for 20 min. Coverslips were then incubated O/N with appropriate primary antibodies and for 1 h with Alexa-594 and Alexa-488 conjugated secondary antibodies (Molecular Probes). Coverslips were mounted on glass slides with Vectashield (Vector Laboratories). Images were taken using a confocal microscope (Leica TCS SP2 AOBS) using a Plan-Neofluar 63× immersion objective (Carl Zeiss, Inc.).

Chromatin immunoprecipitation
ChIP was performed as reported previously (81,82), with minor modifications. Briefly, HeLa TFEB-FLAG stable transfectants (17) were grown up to 50% confluence and cross-linked in 1% formaldehyde for 10 min and lysed on ice for 20 min in ChIP-Lysis buffer (50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1% Triton X-100, 1% Tween-20). After a 13 min MNAse digestion (2 U, Sigma-Aldrich) at 37°C, the reaction was stopped by addition of sodium dodecyl sulfate (SDS) and ethylenediaminetetraacetic acid (EDTA) to a final concentration of 1% and 2 mM, respectively. The unbound SDS of the cleared lysate was precipitated using SDS-OUT (Pierce, Rockford, IL, USA) to avoid compromising the immunoprecipitation. The lysates were diluted 1:1 with ChIP-dilution buffer (50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 0.5% Triton X-100, 2 mM EDTA; all from Sigma-Aldrich) and pre-incubated with high capacity NeutrAvidin Agarose (Pierce). Protein–DNA complexes were immunoprecipitated for 4 h at 4°C with biotinylated FLAG antibody coupled to Neutravidin beads (2 μg ANTI-FLAG BioM2 from Sigma-Aldrich antibody with 50 μl Agarose slurry in ChIP-dilution buffer supplemented with 10 mg/ml BSA per sample). After three washings, the DNA was eluted by addition of 8 mM biotin, 1% SDS in TE buffer. The DNA was precipitated after cross-link reversal using 200 mM NaCl at 65°C, overnight.

Quantitative real-time PCR
Real-time quantitative RT–PCR on digested chromatin was carried out with the LightCycler 480 SYBR Green I mix (Roche) using the Light Cycler 480 II detection system (Roche) with the following conditions: 95°C, 5 min (95°C, 10 s; 60°C, 10 s; 72°C, 15 s), ×40. For expression studies, the qRT–PCR results were normalized against an internal control (APRT). Oligonucleotide sequences are reported in Supplementary Material, Table S5.

Chromatin immunoprecipitation-seq
ChIP samples were sequenced using the standard Solexa protocols. 80 bp reads were mapped to the human genome (UCSC hg19) using the Bowtie algorithm (83), and unique tags were selected for analysis. ChIP-seq areas of enrichment were determined using a custom sliding window approach based on previous work (84) but implemented in C++.

Bioinformatics and data visualization
Annotation, motif analyses, Venn diagrams and pie charts were generated using custom annotation scripts, and statistical significance was assessed with the Wilcoxon rank-sum test or a permutation statistics. Plots of ChIP-seq data relative to RefSeq genes were generated in R by setting all gene lengths equal to 1 and plotting tag density. Statistical and graphical analyses were carried out in the R programming environment.

Genome and expression analyses
Sequence analysis of ChIP peaks was performed with the Weeder webtool (20,21) and the logo of the resulting CLEAR motif was elaborated by using WebLogo (85). Human promoter sequences were retrieved from the Ensemble database (http://www.ensembl.org) and analyzed with the CLEAR PWM by using PatSer from the Regulatory Sequence Analysis Tool package with default parameters (86). A score was assigned to each human promoter, equal to the PatSer score associated with sequences similar to the CLEAR PWM. For promoters with N detected CLEAR-like sequences, the total score was calculated as $S_{TOT} = (S_1^2 + S_2^2 + \ldots + S_N^2)^{1/2}$. GO analyses were performed with the web tool DAVID (87) using default parameters. Redundant terms were manually removed from the resulting lists. Expression correlation analysis was performed as described previously (17,88,89), with minor modifications. Briefly, lysosomal genes were analyzed by using the g:Sorter tool, which is part of the g:Profiler package (31). For a selected gene probe, g:Sorter can retrieve a number of most similar coexpressed profiles in a specified GEO data set. The analysis was carried out on a total of 106 heterogeneous microarray experiments, based on the Affymetrix HG-U133 Plus 2.0 array. g:Sorter was queried with the gene probes for the set of ‘core’ CLEAR genes (Fig. 3) or for a set of established lysosomal genes (Supplementary Material, Table S3). For each analyzed probe, the first 3% of most correlated gene probes was retrieved for each microarray data set. Subsequently, all HG-U133 Plus 2.0 gene probes were ranked based on their cumulative occurrence in the 106 different lists of most correlated genes. The procedure resulted in lists of gene probes ranked by their expression correlation to the investigated lysosomal genes or the subset of ‘core’ CLEAR genes. Homology search and protein domain analysis was performed using the BLAST, BLAT and Swiss-Prot webtools as reported previously (90).

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
We thank A.M. Bianco, S. Briesemeister, C. Di Malta and T. Klisch for technical support, S. Arbucci (Integrated
Microscopy Facility of the Institute of Genetics and Biophysics ‘A. Buzzati-Traverso’, Naples and R.C. Atkinson (Baylor College of Medicine, Houston) for help in the acquisition of confocal microscope images, and A. De Matteis, G. Diez-Roux, A. Flora and T. Klisch for the critical reading of the manuscript.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by the Italian Telethon Foundation; the Beyond Batten Disease Foundation; and the European Research Council Advanced Investigator (grant number 250154 to A.B.).

REFERENCES


Atoh1 disrupts Sonic Hedgehog signaling in the developing cerebellum 
and prevents medulloblastoma. Science, 326, 1424–1427.
vivo Atoh1 targetome reveals how a proneural transcription factor regulates 
and memory-efficient alignment of short DNA sequences to the human 
84. Fejes, A.P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M. and 
from massively parallel short-read sequencing technology. 
Bioinformatics, 24, 1729–1730.
86. Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., 
integrative analysis of large gene lists using DAVID bioinformatics 
88. Gennarino, V.A., Sardiello, M., Avellino, R., Meola, N., Maselli, V., 
prediction by expression analysis of host genes. Genome Res., 19, 
481–490.
89. Gennarino, V.A., Sardiello, M., Mutarelli, M., Dharmalingam, G., 
unique resource for microRNA target prediction. Gene, 480, 
51–58.
90. Sardiello, M., Cairo, S., Fontanella, B., Ballabio, A. and Meroni, G. 
(2008) Genomic analysis of the TRIM family reveals two groups 
of genes with distinct evolutionary properties. BMC Evol. Biol., 
8, 225.