POST-TRANSLATIONAL MODIFICATIONS BY SUMO AND UBIQUITIN-LIKE PROTEINS; ROLE OF SUMOYLATION ON SALL PROTEINS

Lucia Pirone

January 2016

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Facultad de Ciencia y Tecnología

Departamento de Genética, Antropología Física y Fisiología Animal

POST-TRANSLATIONAL MODIFICATIONS BY SUMO AND UBIQUITIN-LIKE PROTEINS; ROLE OF SUMOYLATION ON SALL PROTEINS

Memoria presentada por

Lucia Pirone

Enero 2016

Este trabajo ha sido realizado en la unidad de Genómica Funcional del Centro de Investigación Cooperativa en Biociencias (CIC bioGUNE), financiado con una beca doctoral del proyecto UPStream: European Research Training in the Ubiquitin Proteasome System (Marie Curie Initial Training Network, ITN, FP7A-PEOPLE 2011-ITN) bajo la dirección de la Doctora María Rosa Barrio Olano

CIC bioGUNE
Biozientzietako Ikerkuntza Kooperatiboko Zentroa
Centro de Investigación Cooperativa en Biociencias

MARIE CURIE ACTIONS

INDEX

ABBREVIATIONS

20E	20-hydroxyecdysone
Atg	Autophagy-related Ubiquitin-like modifier
BSA	Bovine serum albumin
CBX4	Chromobox 4
ConA	Concanavalin A
CTBP1	C-terminal binding protein 1
CTCF	CCCTC-binding factor
DAPI	4',6-diamidino-2-phenylindole
DeSI	Desumoylation isopeptidase
Dmnt3	DNA methyltransferase
Dpp	Decapentaplegic
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF4E	Eukaryotic initiation factor 4E
ESC	Embryonic stem cells
FAT10	HLA-F-adjacent transcript 10
Fax	Failed axon connections
FL	Full lenght
FT	Flow Trough
Ftz-f1	Fushi tarazu factor 1
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GO	Gene Ontology
Gro	Groucho
HCSM	Hydrophobic cluster SUMOylation motif
HEK 293FT	Human embryonic kidney 293 FT
I	Isoleucine
ISG15	Interferone-stimulate gene 14
kuk	Kugelkern
L	Leucine
lwr	Lesswright
NEDD8	Neural precursor cell expressed developmentally down-regulated 8
NPC	Nuclear pore complex

NuRD	Nucleosome Remodeling Deacetylase
OS	Okihiro syndrome
Ote	Otefin
PARP	Poly [ADP-ribose] polymerase 1
PBS	Phosphate buffer saline
PBS-T	PBS tween
PBT	PBS, TritonX-100
Pc	Polycomb
PcG	Polycomb group
PDSM	Phosphorylation-dependent SUMOylation motif
PG	Prothoracic gland
PIAS	Protein inhibitor of activated STAT
PLA	Proximity ligation assay
PML	Promyelocytic Leukemia
PMT	Post-translational modification
PRC1	Polycomb repressive complex 1
RanBP2	Ran-binding protein 2
RanGAP1	Ran GTPase activating protein 1
Rel	Relish
RNF4	RING finger protein 4
SAE1	SUMO-activating enzyme subunit 1
SAE2	SUMO-activating enzyme subunit 2
SALL	Spalt-like protein
Salm	Spalt major
Salr	Spalt-related
SDS	Sodium dodecyl sulphate
SENP	SUMO specific peptidase
SF-1	Steroidogenic factor 1
SIM	SUMO interaction motif
SIRT1	Sirtuin 1
SUMO	Small Ubiquitin-like modifier
TBS	Townes-Brock syndrome
Ttk69	Tramtrack69
Ub	Ubiquitin

UBA2	Ubiquitin-like modifier activating enzyme 2
UBA5	Ubiquitin-like modifier activating enzyme 5
UBE2I	Ubiquitin conjugating enzyme E2I
UbL	Ubiquitin-like modifier
Ubx	Ultrabithorax
UFC1	Ubiquitin-fold modifier conjugating enzyme 1
UFL1	UFM1-specific ligase 1
UFM1	Ubiquitin fold modifier 1
UFSP	UFM1-specific peptidase
URM1	Ubiquitin-related modifier 1
Usp	Ultraspiracle
USPL1	Ubiquitin specific peptidase like 1
V	Valine
Velo	Veloren
WB	Washing Buffer
WT	Wild type
ZF	Zing finger
ZNF451	Zing finger protein 451

FIGURES & TABLES

SUMMARY

Post-translational modification by conjugation of Ubiquitin- and Ubiquitin-like molecules contributes to extend the functionality of eukaryotes proteomes. SUMO (Small Ubiquitin-related Modifier) is added and deleted from many cellular substrates to control activity, localization, and recruitment of other SUMO-recognizing protein complexes. The dynamic nature of this modification and its low abundance in cells make it challenging to study, with susceptibility to deSUMOylases further complicating its analysis. In this work we developed a new methodology to isolate and analyze SUMOylated proteins in cultured cells and in a tissue-specific manner in flies. SUMOylated substrates are labelled by *in vivo* biotinylation, which facilitates their subsequent purification using Neutravidin-based affinity chromatography under stringent conditions and with very low background. Based on the success of bioSUMO in *Drosophila*, we have applied the same technology to mammalian Ubiquitin-like (UbL) proteins, using modular multicistronic vectors previously developed in our laboratory. We present here our current toolbox of vectors for application in mammalian cells and insects to isolate proteins that are modified with UbLs of choice, including SUMOs, Ubiquitin, NEDD8, FAT10, ISG15, UFM1, URM1 and other.

In the second part of this Thesis, we focused our investigation on SALL (Spalt-like) proteins SUMOylation. Those proteins are zinc finger transcription factors conserved from *C. elegans* to mammals. In *Drosophila*, the two paralogs (*salm* and *salr)* control the expression of genes involved in wing and central nervous system development, including cell adhesion and cytoskeletal proteins. In vertebrates, mutations in these genes cause hereditary human diseases as Townes-Brocks and Okihiro syndromes. Human SALL1 and *Drosophila* Salm are SUMOylated *in vitro.* Here we demonstrate their SUMOylation in cells and we investigate the role of a putative SALL E3 ligase, using mammalian cells and *Drosophila* as models.

RESUMEN

La modificaciones post-traduccional debidas a la conjugación de Ubiquitina y de proteínas similares a la Ubiquitina (UbLs) contribuyen a extender y diversificar la funcionalidad de los proteomas eucariotas. SUMO (Small Ubiquitin-related Modifier) se une de manera reversible a muchos substratos celulares para controlar su actividad y su localización, así como el reclutamiento de complejos proteicos que puedan reconocer la molécula de SUMO. La naturaleza dinámica de esta modificación y su escasa abundancia en las células hace que el estudio de la SUMOilación sea a menudo difícil de encarar técnicamente**.** Además, la susceptibilidad de las proteínas modificadas a las deSUMOilasas complica aún más su análisis. En este trabajo hemos desarrollado una nueva herramienta para aislar y analizar proteínas SUMOiladas en cultivo celulares y de manera específica de tejido en *Drosophila*. Mediante esta técnica, los substratos SUMOilados son biotinilados *in vivo*, lo cual facilita su subsiguiente purificación, en condiciones astringentes y con muy poco fondo inespecífico, utilizando cromatografía de afinidad basada en Neutravidin. Tras la aplicación exitosa de la tecnología de bioSUMO en *Drosophila*, aplicamos la misma técnica a otras proteínas similares a Ubiquitina en mamíferos, utilizando para ello vectores multicistrónicos modulares desarrollados previamente en nuestro laboratorio. Presento en esta Tesis la batería de herramientas elaboradas que se puedan utilizar en células de mamífero y de *Drosophila* para aislar proteínas modificadas por una variedad de UbLs, incluidas SUMO, Ubiquitina, NEDD8, FAT10, ISG15, UFM1, URM1 y otras.

En la segunda parte de esta Tesis, nos enfocamos en la investigación sobre la SUMOilacion de las proteínas SALL (Spalt-like). Estas proteínas son factores de trascripción del tipo dedo de zinc, conservadas desde *C. elegans* hasta mamíferos. En *Drosophila*, los dos parálogos (*salm* and *salr)* controlan la expresión de genes involucrados en el desarrollo del ala y del sistema nervioso central, incluyendo moléculas de adhesión y proteínas del citoesqueleto. Mutaciones en estos genes provocan enfermedades humanas hereditarias, tales como el Síndrome de Townes-Brocks y el Síndrome de Okihiro. Tanto el homólogo humano SALL1 y como Salm en *Drosophila* están SUMOiladas *in vitro.* Aquí demostramos su SUMOilación en células e investigamos el papel de una posible E3 ligasa de SALL, utilizando células de mamífero y *Drosophila* como modelos.

I. INTRODUCTION

1. Post-translational modifications

The eukaryotic proteome is modulated by a variety of post-translational modifications (PMTs) that contribute to maintain the cellular homeostasis. These modifications could involve small chemical molecules as it occurs in phosphorylation, acetylation, methylation and sulfation, or proteins could be covalently modified by other proteins. This is the case of the modification by Ubiquitin (Ub), which is the result of sequential reactions catalyzed by several enzymes. Proteins similar to Ub, either in sequence or in their three-dimensional structure, form the family of Ub-like modifiers (UbL), which can modify target proteins is a manner similar to the Ub cycle (Fig. 1). Nearly 20 UbL proteins have been described in mammalian cells, which are conserved among eukaryotes and are object of an increasing interest in investigation. The main members of the UbL family include SUMO1-4 (Small Ub-like Modifier1-4), NEDD8 (Neural precursor cell expressed developmentally down-regulated 8), ISG15 (Interferon-stimulate gene 15), FAT10 (HLA-F-adjacent transcript 10), UFM1 (Ub fold modifier 1), Atg-8 and Atg-12 (autophagy-related Ub-like modifier 8 and 12) and URM1 (Ub-related modifier 1) (reviewed by van der Veen & Ploegh 2012). The work presented in *Results Part I* section of this PhD Thesis is focused on SUMOylation, one of the most studied UbL post-translational modifications, and on UFMylation, one of the less studied UbL modifications.

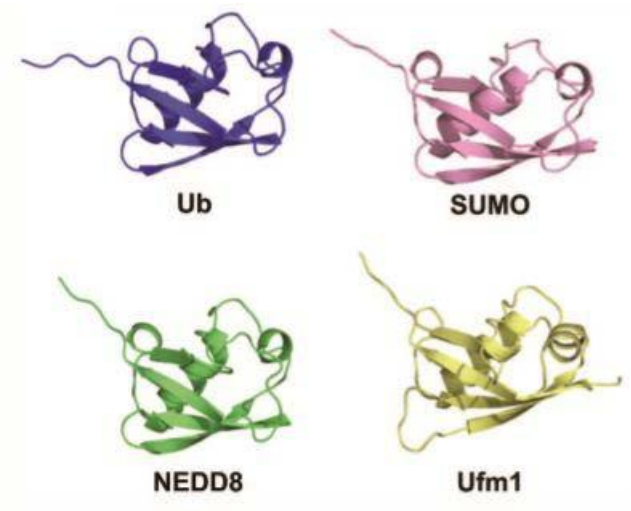


Figure 1. Structure of Ub and UbL proteins. Representation of three-dimensional structure of Ub, SUMO, NEDD8 and Ufm1 proteins that show a common secondary structure. Adapted from Sorokin et al. 2009.

1.1. SUMOylation and SUMO proteins

SUMOylation is a reversible modification involved in diverse cellular and biological processes such as DNA replication, cell cycle progression, transcription regulation, modulation of protein-protein interactions, protein translocation and nuclear trafficking.

At least one SUMO precursor protein was found in all eukaryotes tested since its discovery in 1997 by Mahajan and co-workers (Mahajan et al., 1997). Yeast, *Caenorhabditis elegans* and *Drosophila melanogaster* are among the species expressing only a single SUMO protein; in vertebrates, four SUMO paralogs were identified, denominated SUMO1-4 (reviewed by Flotho & Melchior 2013). Mature human SUMO2 and SUMO3 are very similar (97% sequence identity) and, since they cannot be distinguished by available antibodies, usually both proteins are referred to as a sub-group SUMO2/3. On the other side, SUMO1 diverges from SUMO2/3, sharing only 47% sequence identity. Despite diversity between the two sub-groups, all SUMO proteins are characterized by the Ub fold (globular β-grasp fold) and a C-terminal glycine-glycine motif that is exposed after their maturation. SUMO4 was described in human cells, its sequence being 87% similar to SUMO2 and its maturation being observed in lysates from starved-cells. However, the details of its processing and whether the conjugation to specific targets is actually occurring, are still unclear (Guo et al., 2004; Wei et al., 2008).

Human SUMO paralogs show different ability to form polySUMO chains *in vitro* and *in vivo* (Tatham et al., 2001). SUMO2 and SUMO3 contain an internal consensus site for SUMOylation that allow the covalent linkage by other SUMOs molecules; SUMO1 could be also incorporated in these chains, but the lack of SUMO consensus site limits the elongation after its incorporation (Matic et al., 2008). Formation of polySUMO1 chains has been reported only *in vitro* in presence of an excess of SUMO1 and a target protein (RanBP2, Ran-binding protein 2) (Pichler et al., 2002). In yeast, Smt3 is able to form polySUMO chains regulated by the isopeptidase Ulp2 while, recently, it has been demonstrated that orthologs of human SUMO3 in insects have lost the capability to form polySUMO chain (Bylebyl et al., 2003; Ureña et al., 2015)

1.2. The SUMOylation pathway

Like ubiquitination, the process of SUMOylation consists of an enzymatic cascade involving the activation, conjugation and ligation of the SUMO moiety by E1, E2 and E3

enzymes, respectively (Fig. 2). Despite differences between SUMO2/3 and SUMO1, the same catalytic enzymes activate and conjugate diverse SUMO moieties to target proteins. The case of SUMO4 is different, mainly due to the presence of a proline residue (Pro90) at the C-terminus that blocks the enzymatic reaction of activation. Data are not available and still is not clear how the maturation process occurs for this paralog (Owerbach et al., 2005).

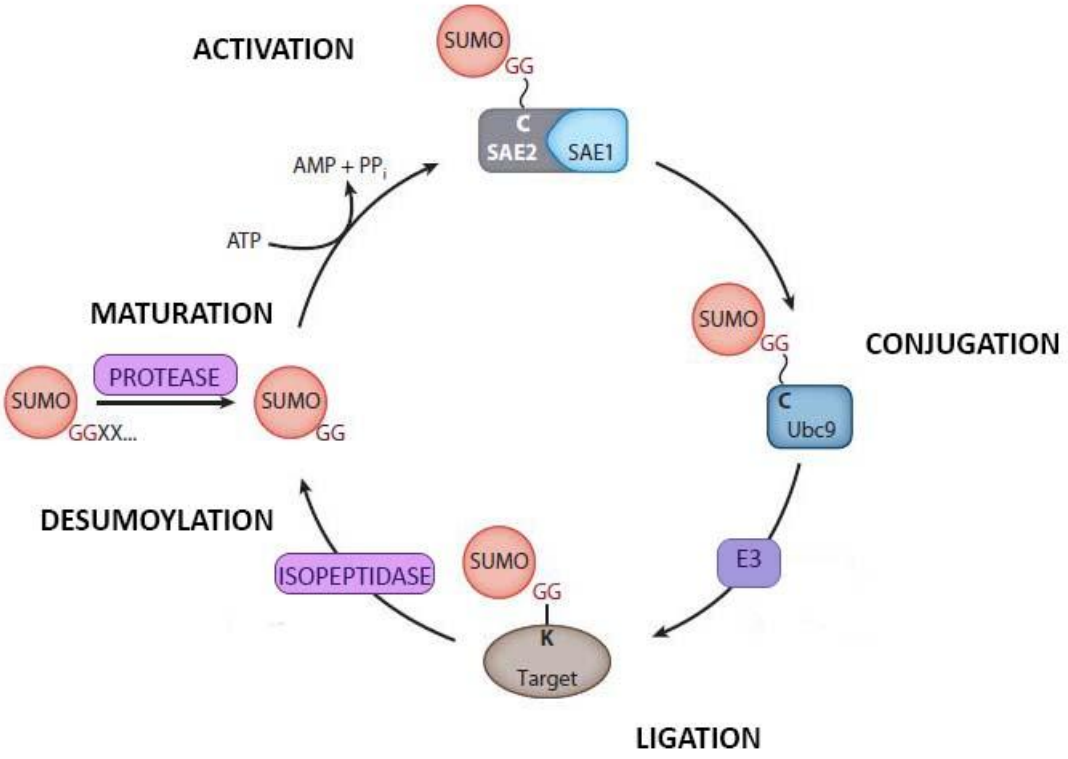


Figure 2. Schematic representation of SUMOylation cycle. Adapted from Flotho & Melchior 2013. See text for details.

Prior to the first conjugation cycle, a nascent SUMO needs to be cleaved by a SUMO-specific protease belonging to the Ulp/SENP family. The elimination of the C-terminal residues allows the exposure of the glycine-glycine motif. In the first step, the heterodimeric E1, composed by SUMO activating enzymes 1 and 2 (SAE1 or AOS1 and SAE2 or UBA2), activates the C-terminus of a mature SUMO moiety by ATP hydrolysis and forms a thioester bond with a cysteine residue in SAE2. In the second step a unique E2, called UBE2I or UBC9, binds to SAE2 and accepts the SUMO moiety from the E1 on its active cysteine by a transthiolation reaction. Finally, an E3 ligase catalyzes the transfer of SUMO from the E2 to a

target protein forming an isopeptidic bond between the ε-amino group of a lysine in the substrate protein and the C-terminal glycine in the SUMO moiety (reviewed by Eifler & Vertegaal 2015).

SAE1 and SAE2 are encoded by unique genes in the human genome. As mentioned, the E2 is represented by one protein, in contrast to the 35 conjugating enzymes involved in the ubiquitination machinery. Despite being the only SUMO E2, UBC9 plays a regulatory role in target protein selection (Knipscheer et al., 2008). Modulation of the UBC9 expression influences the entire SUMOylation process and has been demonstrated that high concentration of E2 in presence of E1 is able to SUMOylate target proteins in absence of E3 ligases (Bernier-Villamor et al., 2002).

Despite this, the E3 SUMO ligases promote SUMOylation under physiological conditions, and confer specificity to the substrate. In contrast with the ubiquitination pathway, only a few SUMO ligases have been described. These are classified in two sub-families: (i) the Siz/PIAS RING ligase sub-family and (ii) the nucleoporin RanBP2 sub-family. Six PIAS (Protein inhibitor of activated STAT) members belong to the first group and are characterized by three conserved motifs: (I) a N-terminal PINIT domain that plays a unique role in substrate recognition, (II) a central zinc finger RING-like domain (SP-RING), which recognizes the cognate E2 into a complex and facilitates the SUMO conjugation, and (III) a C-terminal domain (SP-CTD) required for the interaction with SUMO. Like it occurs during ubiquitination, the RING domain provides the scaffold to bring UBC9 and the substrate together (Yunus and Lima, 2009). The other ligases group is characterized by RanBP2 protein that is part of the nuclear pore complex (NPC) and is localized in cytoplasmic filaments, implying that the SUMOylation process is not exclusive of the nucleus. In contrast to the RING E3 ligases, RanBP2 does not bind to the target protein but enhance its SUMOylation, as demonstrated in the case of Sp100 (Pichler et al., 2002) or Ran GTPase activating protein 1, RanGAP1 (Werner et al., 2012).

Several other proteins have been identified as E3 ligases that could not be classified in these two sub-families. Among these, the human Polycomb (Pc) CBX4/Pc2, component of the multimeric Polycomb repressive complex 1 (PRC1), which enhances the SUMOylation of C-terminal binding protein 1 (CTBP1) (Kagey et al., 2003), the CCCTC-binding factor (CTCF) (MacPherson et al., 2009), or the DNA methyltransferase 3, Dmnt3 (Li et al., 2007). Also, it regulates the recruitment of the Pc complex protein BMI1 at DNA-damaged sites trough SUMOylation (Ismail et al., 2012).

SUMOylation is a dynamic and reversible process. DeSUMOylation is catalyzed by SUMO isopeptidases that are, in some cases, also SUMO proteases, a family of cysteine proteases necessary for the SUMO moiety maturation prior to the activation and conjugation

cycle. The first protein family of SUMO protease described was Sentrin/SUMO specific peptidase (Ulp/SENP). Human genome encodes for SENP1, SENP2, SENP3, SENP5, SENP6 and SENP7, that show conserved catalytic domain localized at the C-terminus of the proteins. The subcellular distribution of isopeptidases contributes to the specificity for their targets: SENP1 and SENP2 are concentrated in the nuclear envelope trough their interaction with the NPC; SENP3 and SENP5 localize in the nucleolus where they participate in the early step of ribosome maturation; SENP6 and SENP7 exhibit a nucleoplasmic distribution. SENP members are also characterized by the different affinity for diverse SUMO moieties: SENP1 shows a mayor affinity for SUMO1 in SUMO processing and SUMO deconjugation; SENP2 catalyzes more efficiently the deconjugation for SUMO2; the couple SENP3-SENP5 acts on SUMO2/3 for processing and deconjugation; and SENP6-SENP7 exhibit preference for deconjugation of polySUMO2/3 chains (reviewed by Nayak & Müller 2014). Other SUMO isopeptidases have been described: the DeSI family composed by DeSI-1 and DeSI-2 that shown only isopeptidase activity; and the Ubiquitin-specific protease-like 1 (USPL1), a Cajal body protein efficient only in SUMO deconjugation with preference for polySUMO2/3 chains (Schulz et al., 2012) . As E3 ligases, SUMO isopeptidase show specificity for their targets (Shin et al., 2012).

1.3. SUMOylation consensus motifs versus SUMO interaction

The identification of hundreds of SUMOylated targets, propitiated by the development of Mass Spectrometry strategies to identify the SUMO insertion sites, contributed to demonstrate that the interaction of the SUMO moiety occurs on a lysine. Based on experimental evidences, and in contrast with ubiquitination, it was possible to develop bioinformatics tools to predict SUMOylation consensus sites as SUMOplot program or GPS-SUMO (Zhao et al., 2014). In addition, SUMO proteins could also interact with proteins in a non-covalent way trough specific recognition sequences called SUMO Interaction Motifs or SIMs. SUMO consensus motifs and SIMs will be described more in detail in the next paragraphs to distinguish between SUMO conjugation and SUMO interaction.

1.3.1. SUMO consensus motif

SUMOylation targets are modified by the covalent attachment of SUMO to a lysine generally present in the canonical SUMO consensus motif ΨKxD/E, where Ψ is a hydrophobic residue, K is the conjugation lysine, x is any amino acid, and D or E are aspartic or glutamic acids, respectively (Rodriguez et al., 2001). Other sequences, different from the canonical consensus motif, were described in several studies: an inverted consensus motif D/ExKΨ (Matic et al., 2010); an hydrophobic cluster SUMOylation motif (HCSM), consistent of an extended consensus motif where the target lysine is preceded by a cluster of at least three hydrophobic amino acids (Matic et al., 2010) and a phosphorylation-dependent SUMOylation motif (PDSM), composed of a SUMO consensus site and an adjacent proline-directed phosphorylation site (Hietakangas et al., 2006). A large-scale site-specific analysis of SUMOylation sites suggests that ΨKxE could be a better substrate than ΨKxD for sites with the forward consensus motif, while in the inverted consensus motif the presence of aspartic or glutamic acids was found in the same proportion in the identified targets (Tammsalu et al., 2014).

1.3.2. SUMO non-covalent interactions

The first evidence of a specific sequence that mediates a non-covalent binding with SUMO was reported in 2000, when a short stretch of branched hydrophobic residues forming a SIM was identified by yeast two-hybrid screen (Minty et al., 2000). Few years later, the presence of three residues (valine, V, leucine, L, or isoleucine, I) was reported as essential. These residues could be arranged as V/I-x-V/I-V/I or V/I-V/I-x-V/I/L and could bind SUMO in the two orientations (Song, 2005). SIM domains interact with the SUMO moieties in a very specific way. Three-dimensional structural studies revealed that this motif could be consider as a “code of specificity” for SUMO isoforms: the affinity for SUMO1 depends on the presence of negative charges close to the SIM, while the hydrophobic core of the SIM stabilizes the interaction with SUMO2 without necessity of additional charge (Hecker et al., 2006).

The interaction SUMO-SIM has been identified in numerous SUMO targets and enzymes, being involved in SUMOylation and in the regulation of protein-protein interactions. Non-covalent interaction between SUMOylated UBC9 and its targets trough SIMs stimulates proteins SUMOylation and promotes chain formation (Knipscheer et al., 2007). SIMs are present in various E3 ligases: RanBP2 catalytic activity depends on its SIM; PIASα is phosphorylated

within its SIM, which influences its binding to SUMO; CBX4 SUMOylation depends on its SIMs, necessary also for its E3 ligase activity (reviewed by Flotho & Melchior 2013; Kagey et al. 2005; Merrill et al. 2010).

A different role for SIM motifs was identified for the RING finger protein 4 (RNF4), an Ubiquitin E3 ligase that recognizes, through its SIMs, the polySUMO chains of Promyelocytic Leukemia (PML) and targets the SUMOylated protein for degradation mediated by ubiquitination. This represents an example of a cellular mechanism that involves a cross-talk between two post-translational modification (Tatham et al. 2008, Xu et al. 2014).

To conclude this paragraph on SIM motifs, it should be mentioned the recent characterization of ZNF451, a representative member of a new vertebrate family of proteins that are E3 ligases showing a new E4 elongase activity. This consist on the capability to extend a SUMO chain by adding SUMO2/3 moieties, while showing an inefficient initial conjugation of the modifier to the target (Eisenhardt et al. 2015; Cappadocia et al. 2015),

1.4. Strategies to study SUMO modification

Multiple strategies have been used to study the post-translational modification of target proteins. Despite the increasing number of SUMOylated targets identified during the last years, the analysis of SUMOylated proteins is still challenging compare to the study of other PTM as ubiquitination or phosphorylation. The main reason is that the general SUMO expression level is lower than the one exhibited by Ub, and this is reflected on the small proportion of SUMOylated fraction in the total pool of a given protein. In addition, the susceptibility to deSUMOylases further complicates the isolation of SUMOylated proteins. In the next paragraphs I will briefly describe the advantages and disadvantages of the different approaches available for the detection of SUMOylated proteins.

An efficient SUMO target enrichment is achieved by the fusion of the N-terminus of SUMO to a tag, single or in tandem repeats, like for example His_6, $His_{10,}$ FLAG, Myc, His_6-FLAG, His_6-HA, FLAG-TEV or ProtA-TEV-CBP, which could be expressed transient or stably in cell lines, as well as in transgenic organisms (Denison, 2005; Ganesan et al., 2007; Hannich et al., 2005; Nie et al., 2009; Schimmel et al., 2014; Tirard et al., 2012; Vertegaal et al., 2004; Wohlschlegel et al., 2004). Depending on the tag, the conditions used for purification of the modified proteins could be more or less stringent, which is important to inactivate deSUMOylase activities. For example, in the case of His_6 or $His_{10,}$ guanidine or urea buffers are conveniently

used but, on the other hand, endogenous proteins that naturally contain histidine rich sequences are also purified with the nickel chromatography. These N-terminal fusions require the exogenous expression of SUMO, which could not reflect the endogenous conditions in terms of expression levels and, consequently, could influence global SUMOylation. A very important advantage of those SUMO fusions is the possibility to mutate SUMO isoforms by point mutations as Q87R, T90R and T90K, which generates an additional cleavage site near the C-terminus to facilitate Mass Spectrometry analysis (Tammsalu et al., 2014). This approach is very useful for the identification of SUMO target lysines, although it is necessary to verify that the functionality of the mutant SUMO, in terms of conjugation and deconjugation, reflects the wild type (WT) one. Recently, a method for identification of SUMOylation sites without using SUMO variants has been developed, based on the chemical blocking of all free lysines, followed by treatment with SUMO specific proteases and subsequent identification of the 'freed' lysines by high-resolution Mass Spectrometry (Hendriks et al., 2015).

The best hypothetical approach is the identification of modified targets under completely endogenous condition. In 2013, Becker and co-workers identified 600 SUMO targets by using monoclonal antibodies against the glycine-glycine C-terminal dimer (Becker et al., 2013). The biggest advantage of this method is the possibility to use these antibodies for any type of sample, including primary cells that are hard to transfect for exogenous expression or rare patient material. However, this method is relatively expensive and requires large amounts of sample material, as well as a complex purification protocol. Finally, SUMO-traps have been developed to isolate endogenous modified proteins. In this case, the affinity is very high for poly- or multi-SUMOylated proteins and requires as well large amounts of sample material (Bruderer et al. 2011, Da Silva-Ferrada et al. 2013).

1.5. The SUMOylation process in *Drosophila*

1.5.1. *Drosophila melanogaster* as a model system

The fruit fly *Drosophila melanogaster* is a versatile model organism that has been used over a century for genetic and developmental studies. There are many technical advantages for using *Drosophila* over vertebrate models: it is easy and inexpensive to culture in laboratory conditions, there are numerous established techniques for genetic manipulation, the life cycle is relatively short (11 days at 25°C) and the offspring production is quite large.

Drosophila life cycle consists in four morphologically distinct stages. Once fertilized, the embryo develops from the egg and molts into larval stages after 24 hours at 25°C. The larval stages consist of three instars, LI, LII and LIII, over five days, where most of the body growth occurs and the animal attains its final size. After the larval stages, the animal enters into pupariation for 4-5 days, during which metamorphosis occurs. Finally, the adult fly emerges from the pupa. All these developmental transitions are regulated by peaks of steroid hormones necessary to pass from one stage to another (Fig. 3) (Dubrovsky, 2005). A series of enzymatic steps within the endocrine organ of the insect, the prothoracic gland (PG), converts cholesterol into 20-hydroxyecdysone (20E), the most important steroid hormone in insects. 20E is then liberated into the hemolymph and reaches the target tissues.

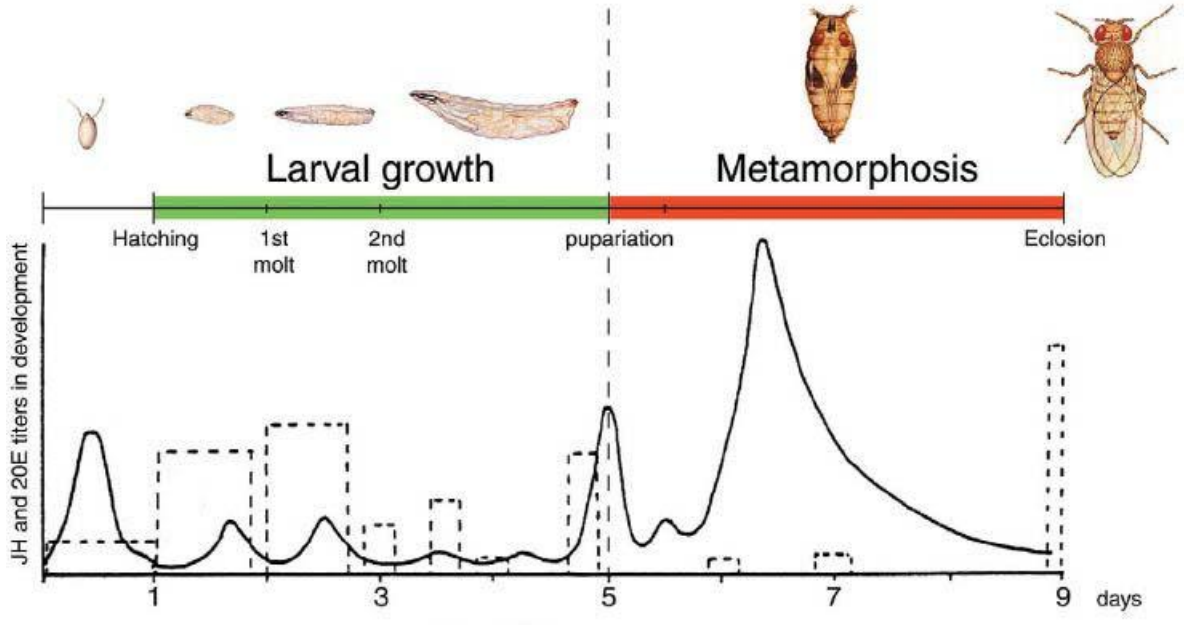


Figure 3. Levels of 20E and juvenile hormone (JH) during *Drosophila* development. Schematic representation of *Drosophila* life cycle with indication of the number of days for each stage**.** A dotted line indicates levels of juvenile hormone and a continue line represents levels of 20E. Developmental transitions are determined by an increase of 20E. Adapted from Dubrovsky 2005.

During metamorphosis, most of the larval tissues are destroyed. Most of the adult tissues such are wings, legs or eyes, develop from the imaginal discs. These are groups of epithelial cells that will form the epidermal structures in the adult. The imaginal discs are present from early embryonic stages, growing in size during larval stages. During the pupal stage they evert and finally develop into the adult structures.

1.5.2. SUMOylation pathway in *Drosophila*

Yeast, *Drosophila* and nematodes genomes encode for a single SUMO gene. *Drosophila smt3* shares 52% and 73% sequence identity with human SUMO1 and SUMO2, respectively. *smt3* is expressed during development, principally during embryogenesis, imaginal discs and in adult female (Ohsako & Takamatsu 1999; Long & Griffith 2000; Lehembre et al. 2000; Kanakousaki & Gibson 2012). The factors involved in the conjugation pathway are also highly conserved in eukaryotes (Talamillo et al. 2008). A component of Ulp/SENP family is involved in Smt3 maturation: after elimination of just two amino acids, *Drosophila* Smt3 enters into the three steps process of activation and conjugation to the target proteins, as described before for vertebrates. In *Drosophila* the activating enzyme E1 is an heterodimer composed by two proteins Aos1 and Uba2 (Long & Griffith 2000; Bhaskar et al. 2000). The unique E2 homolog of UBC9 is encoded by a gene called *lesswright* (*lwr*) (Lehembre et al., 2000), and the E3 ligases identified, Su(var)2-10 and Tonalli, contain a SP-RING as the PIAS family (Hari et al. 2001; Monribot-Villanueva et al. 2013). The deconjugation enzymes characterized are: Ulp1, a protease associated to the NPC responsible of the maintenance of SUMOylated proteins into the nucleus (Smith et al., 2004); and Veloren (Velo), an isopeptidase directly involved in deSUMOylation of Pc (Gonzalez et al., 2014) and with a role in regulation of neuron projection (Berdnik et al., 2012).

The general localization of Smt3 is nuclear and, like the mammalian homologs, *Drosophila* Smt3 localizes in sub-nuclear bodies. The nature of these bodies is still unclear. In mammalian cells, SUMO localizes in PODs (Promyelocytic Leukemia, PML, oncogenic domains) that contain SUMO-conjugated proteins. However, *Drosophila* does not encode for a homolog of PML, therefore the nuclear bodies might contain other proteins that are not yet identified.

1.5.3. The role of SUMOylation in *Drosophila*

SUMOylation in *Drosophila* has been involved in embryogenesis (Epps and Tanda, 1998), growth and proliferation of the imaginal discs (Kanakousaki & Gibson Development 2012), development of the nervous system (Badenhorst et al. 2002; Lehembre et al. 2000), wing morphogenesis (Takanaka and Courey, 2005), metamorphosis (Talamillo et al. 2008) and the immune response (Bhaskar et al., 2002).

Alterations in the expression of *smt3* or of the SUMOylation machinery cause dramatic developmental defects and early lethality due to the regulation of proteins required in the early *Drosophila* embryo (Nie et al., 2009). For example, *smt3* transcripts reduction in embryos is lethal before the second larval instar (Nie et al., 2009); the E2 *lwr* mutation *semushi* is lethal at late embryonic stages, as the nuclear import of Bicoid (a morphogenetic protein that forms a concentration gradient along the anterior-posterior axis) is blocked during the early embryogenesis (Epps and Tanda, 1998); the consequence of *Uba2* overexpression is an increase of Smt3-conjugated proteins that is associated to lethality when occurs in neurons (Long and Griffith, 2000); silencing *smt3* expression in the PG blocks development at the end of LIII due to reduction in the production of 20E, which does not reach the levels necessary to entry into pupariation (Talamillo et al. 2008).

Numerous SUMOylation target have been identified in *Drosophila*. Most of them are transcription factor, as for example p53 that has a fundamental role in the activation of apoptosis; the transcriptional corepressor Groucho (Gro), expressed in the embryo and in the imaginal discs, that regulates several processes from embryonic patterning to neurogenesis; the transcriptional repressor Tramtrack69 (Ttk69) that antagonizes neuronal fate determination in the peripheral nervous system; Dorsal and STAT92E implicated in the immune response; or the zinc finger transcription factors Spalt major (Salm) and Spalt-related (Salr) that control growth and vein formation in wing morphogenesis (reviewed by Smith et al. 2012). We will talk more extensively about the Salm and Salr factors in the paragraphs below.

1.6. UFM1 and the UFMylation pathway

UFM1 is a 9.1 KDa protein with a low sequence identity with Ub but with a very similar tertiary structure. In contrast with Ub and other UbLs, Ufm1 possesses a single active glycine at the C-terminus, being synthesized as an inactive precursor form that has 2 additional amino acids beyond the conserved glycine. Like described above for SUMOylation, UFM1 conjugation utilizes a three-step enzyme system: UBA5 as activating enzyme, UFC1 as E2 conjugating enzyme and UFL1 as an E3 ligase (Fig. 4). Modification of proteins with UFM1 is also reversible. Two novel cysteine proteases have been identified to date (UFSP1 and UFSP2) which cleave UFM1-peptide C-terminal fusions and also remove UFM1 from native intracellular conjugates. These proteases have no obvious homology to Ub deconjugating enzymes.

UFM1 system is conserved in animals and plant, but not in yeast, suggesting an important role in multicellular organism. The role of UFM1 modification is still not completely understood. Some of the identified targets are involved in the endoplasmic reticulum stress response (Lemaire et al., 2011), erythrocyte differentiation in mice (Tatsumi et al., 2011) or in human breast cancer (Zhang et al., 2012). Being one of the most unknown UbLs, its study represents an interesting research challenge.

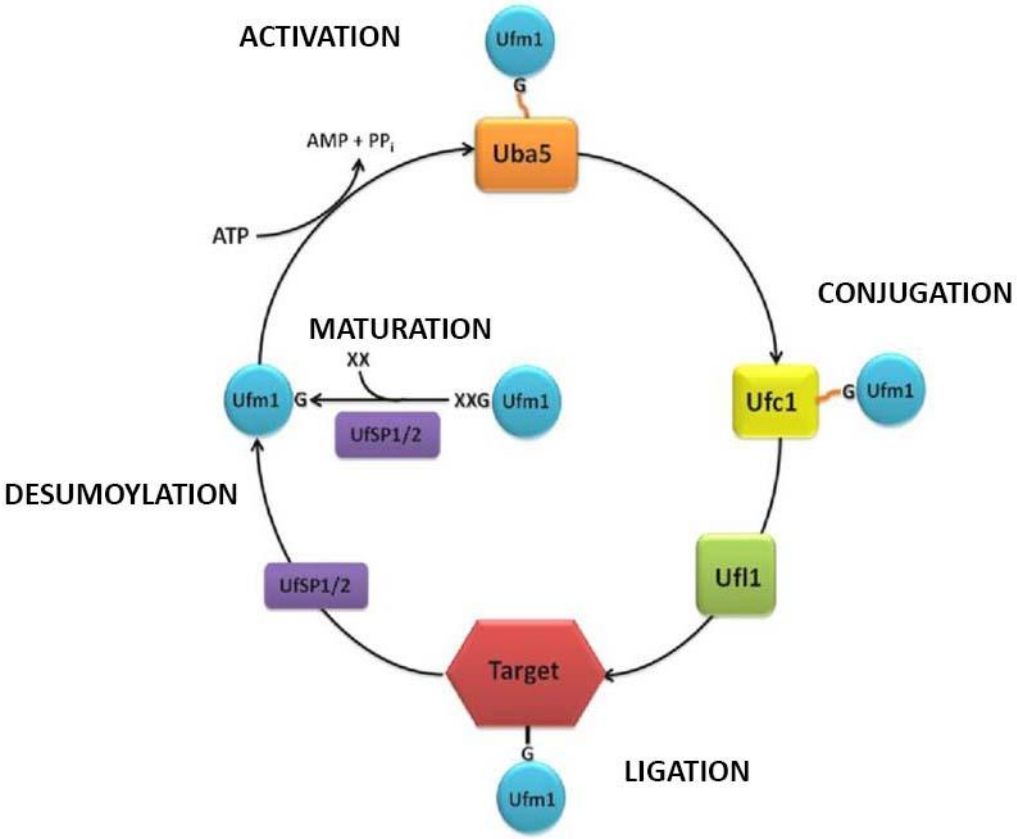


Figure 4. Schematic representation of UFMylation cycle. Adapted from Daniel & Liebau 2014. See text for details.

2. The SALL family of transcription factors

SALL (Spalt-like) proteins are zinc finger transcription factors conserved throughout evolution from *C. elegans* to mammals. *SALL* genes were identified for the first time in *Drosophila*, where two paralogs Salm and Salr are encoded (Kühnlein et al. 1994; Barrio et al. 1996; de Celis & Barrio 2009). In *C. elegans* the only SALL gene described is *sem-4*, three genes were identified in zebrafish and in chicken (Farrell et al. 2001; Camp et al. 2003;

Barembaum & Bronner-Fraser 2004; de Celis & Barrio 2009), while the genome of mouse and human encode for four *SALL* genes (*SALL1-4*) (Buck et al. 2001; Kohlhase et al. 1998).

In *Drosophila melanogaster sall* genes were identified as homeotic genes expressed during embryogenesis in the nervous and tracheal systems, as well as in the larval imaginal discs (Kühnlein et al. 1994; reviewed by de Celis & Barrio 2009). The contribution of Sall factors is essential for tracheal and nervous system development (Kühnlein & Schuh 1996; Cantera et al. 2002), for the determination of wing patterning (De Celis and Barrio, 2000) and for sensory organ development (Barrio et al. 1999; Rusten et al. 2001). In *C. elegans*, *sem-4* is implicated in fate determination of different cell types, as neurons, muscles and hypodermis (Toker et al., 2003). Finally, in vertebrates, SALL proteins are required for the correct development of limbs and nervous system, as well as for organs such heart and kidneys.

As shown in Fig. 5 SALL proteins are characterized by the presence of zinc fingers with a precise pattern along the proteins. The first zinc finger domain (ZF1) corresponds to a single zinc finger C_2HC type conserved only in vertebrates; the rest of the domains (ZF2-5) are organized in doublets or triplets of C_2H_2 type zinc fingers, connected by sequences conserved throughout evolution. The N-terminal part of SALL proteins contains a glutamine rich region conserved in vertebrates and invertebrates, which could have a role in protein dimerization or interaction with other proteins. (Kohlhase et al. 1998; Buck et al. 2000; Borozdin et al. 2006). In chickens, this domain is required for the interaction between SALL1 and SALL3 or with other SALL proteins. However, it is unclear whether the dimerization of the two SALL proteins occurs only trough the glutamine rich motif or the binding could be directed by other motifs or mediated through unrelated proteins (Sweetman et al., 2003).

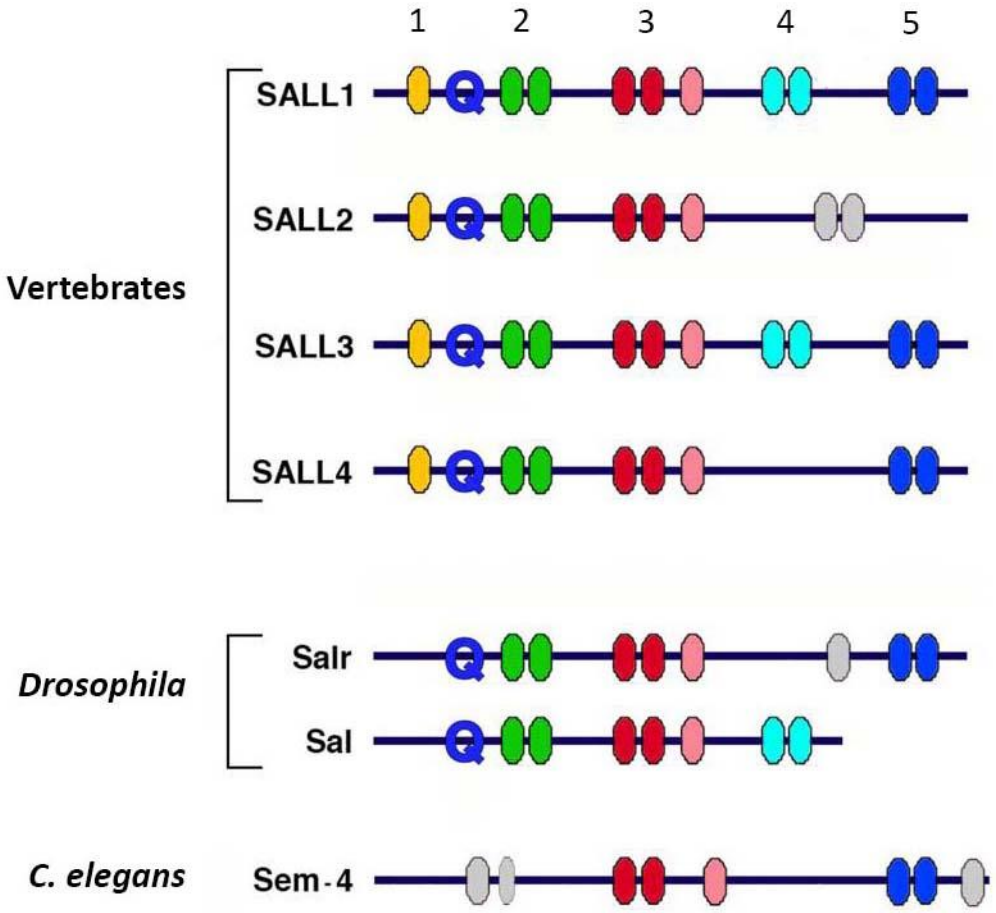


Figure 5. Schematic representation of SALL proteins in vertebrates, *Drosophila* and *C. elegans*. Colored ovals indicate the zinc finger domains, the number of each domain being indicated in the top of the figure. Q represents the glutamine-rich region. Adapted from de Celis & Barrio 2009.

The two *Drosophila* SALL paralogs have transcriptional repression activity acting through AT-rich DNA sequences (Sánchez et al. 2011; Barrio et al. 1996). Human SALL1 and SALL3 are classified also as transcriptional repressors, while SALL2 and SALL4 activate transcription, as described in mouse and in human trough CT- or GC-rich sequences. Two possible repression mechanisms have been described for SALL1: a conserved sequence composed by 12 amino acids at the N-terminus of the proteins is responsible for the Nucleosome Remodeling Deacetylase (NuRD) complex recruitment, one of the major corepressor complexes in mammalian cells, and its associated histone deacetylase activity (Lauberth and Rauchman, 2006); the second mechanism requires the central region of the protein including ZF2 and ZF3 (Netzer et al., 2006).

2.1. SALL proteins in human health

The SALL family is very important in different aspects of human health, since they are associated to hereditary syndromes and involved in stem cell maintenance and cancer as tumor-suppressor factors. Mutations in *SALL1* and *SALL4* genes cause the Townes-Brocks (TBS) and the Okihiro syndromes (OS), respectively.

Figure 6. Clinical features of TBS and OS syndromes. Patients affected by either of these two syndromes exhibit hand (polydactyly) and foot malformations, as well as external ear malformations. Pictures are taken form a patient suffering TBS. Adapted from Kohlhase et al. 1998.

TBS is an autosomal dominant syndrome characterized by thumb abnormalities, dysplastic ears and imperforated anus. Additional malformations have been described in patients with TBS including structural and functional renal anomalies, hand malformations, foot malformations, hearing loss, congenital heart defects, and eye anomalies.

Since its discovery, many familiar and isolated cases have been described. More than 60 point-mutations were described that cause premature stop codons by frame shifts, short insertions or deletions, mainly in a hot-spot region located between the N-terminal part of the protein and ZF2 (Kohlhase et al. 1998; Botzenhart et al. 2005; Botzenhart et al. 2007). In addition, three different deletions, including most part of the SALL1 coding region, have been reported in TBS patients (Borozdin et al. 2006; Miller et al. 2011).

Haploinsufficiency was the first hypothesis to explain the symptoms of patients affected by TBS. However, deletion of murine *Sall1* in heterozygous animals did not show the characteristic TBS phenotype, while in homozygosis only hypoplastic kidneys were observed (Nishinakamura et al., 2001). Further and independent studies suggested that truncated SALL1 results into a dominant negative molecule interfering with the role of WT SALL1 and with its nuclear localization (Botzenhart et al., 2007).

OS is characterized by hearing dysfunction, external ear malformation, renal abnormalities, atrial septal defects and facial asymmetry. Mutations and deletion in the gene *SALL4* are directly connected with the symptoms observed in OS patients. Unlike *SALL1* mutations in TBS, the mutations observed in OS cases do not cluster around a critical region of *SALL4* and haploinsufficiency has been shown to be the cause of this disorder (Kohlhase et al. 2002; Borozdin et al. 2004; Borozdin et al. 2007).

SALL2 and *SALL3* have not been traditionally associated to genetic syndromes. Only recently, mutations in *SALL2* have been identified as responsible for ocular coloboma in human and in mice, a congenital defect resulting from failure in the optic fissure normal closure (Kelberman et al., 2014).

In recent years, a very important role is assigned to SALL proteins in stemness and cancer biology. SALL4 is critical for maintaining pluripotency in mouse embryonic and hemapoietic stem cells (ESC) (Sweetman & Münsterberg 2006; Wu et al. 2006; Yang et al. 2010; Xiong 2014), showing also a role in DNA damage response in ESCs by interaction with the epigenetic machinery (Xiong et al., 2015). SALL2 has been described as a suppressor of ovarian cancer (Sung et al., 2013). Reduced levels of expression of *SALL2* are associated to different tumors as lung, colorectal, breast and prostate cancers, while high levels of the protein are detected in Willm's Tumors (Ma et al. 2001; Li et al. 2002; Liu et al. 2007). In respect to SALL3, it was shown that its methylation is significantly increased in bladder cancer (Yu et al., 2007) and a reduction of *SALL3* expression is observed in human hepatocellular carcinoma (Yang et al. 2012).

2.2. Post-translational modification of SALL proteins

Regulation and localization of transcription factors activity could be modulated by post-translation modifications. In *Drosophila*, SALL proteins can be SUMOylated. This modification alters their nuclear localization and, in addition, their SUMOylation state influences their role in vein pattern formation in the wing (Sánchez et al., 2010), as well as their transcriptional repressor activity (Sánchez et al 2011). Phosphorylation of Sall1 by protein kinase C modifies its transcriptional repression activity in *Xenopus laevis* (Lauberth et al., 2007). Yeast-two-hybrid screen performed using a human library identified UBC9 and SUMO1 as interactors of SALL1 (Netzer et al., 2011). The binding between the E2 and SALL1 and the conjugation of

SUMO1 to the lysine 1086 of SALL1 were confirmed by *in vitro* assays (Netzer et al., 2002) (Fig. 7).

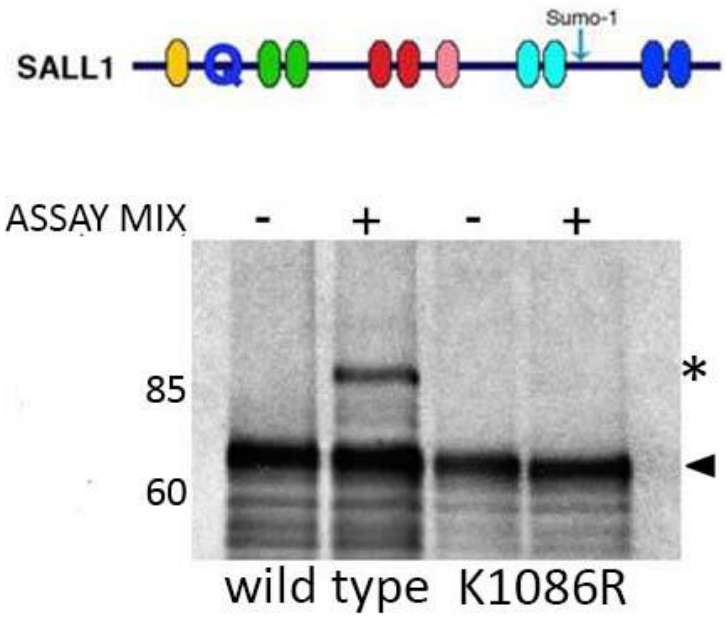


Figure 7. SALL1 SUMOylation *in vitro*. Above, schematic representation of human SALL1, where a potential SUMOylation site is indicated. Below, Western blot showing the result of an *in vitro* SUMOylation assay. A fragment of SALL1 (aminoacids 689-1324) was translated *in vitro* and incubated in absence (-) or presence (+) of an assay mix containing SUMO1, UBC9 and AOS1/UBA2. SUMOylated SALL1 is indicated by an asterisk, while the arrowhead indicates the unmodified protein. Substitution of lysine 1086 for an arginine abolished the conjugation of SUMO1 to this fragment of SALL1. Adapted from Netzer et al. 2002.

SALL4B, the major splicing variant of *SALL4*, is phosphorylated, ubiquitinated and SUMOylated. SUMOylation occurs on four lysines and is important for the stabilization and localization of the protein (Yang et al. 2012).

2.3. Role of SALL proteins during wing development in *Drosophila*

Drosophila salm and *salr* are the main effectors downstream of Decapentaplegic (Dpp), one of the most important morphogens necessary for the correct development of many tissues, Dpp signaling pathway being required for growth and patterning of the fifteen imaginal discs during larval development.

2.3.1. Development of the wing imaginal disc

Most of external structures of the *Drosophila* body develop from the imaginal discs that, as mentioned previously, are larval epithelial structures originated by proliferation of cells localized in the embryonic ectoderm (Bate and Arias, 1991). Growth and patterning of the imaginal discs take place during larval stages. Although the cells in the growing discs appear undifferentiated, their developmental fate is already determined. During pupal stages the discs evert and elongate, the central portion of the disc becoming the distal part of the corresponding appendage (wing, leg, antenna, etc.).

Each wing imaginal disc differentiates into one wing proper and one hemithorax during pupal development. Multiple signaling pathways are involved in wing disc development, including Engrailed (En), Hedgehog, Notch, Wingless, Epidermal growth factor receptor and Dpp signaling pathways. They stabilize morphogenetic gradients that contribute to define the wing disc blade and to subdivide the wing disc along the proximodistal axis. Differential expression of *en* and *apterous* defines the posterior and the dorsal compartments. Dpp signaling, during pupal stages, directs cells towards vein formation.

Drosophila wings have four complete longitudinal veins (LII–LV) distributed over the length of the wing, two transverse veins (anterior and posterior cross-veins; a-cv and p-cv) that connect the longitudinal LIII–LIV and LIV–LV, respectively, and two incomplete veins (LI and LVI) (rewieved by de Celis 2003) (Fig.8).

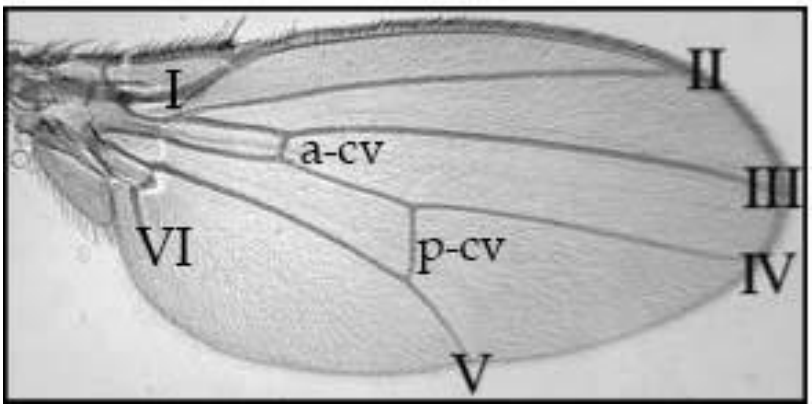


Figure 8. *Drosophila* vein pattern in wing. WT adult wing showing four complete veins LII-LV, two incomplete veins LI and LVI and the two transverse veins a-cv and p-cv. Adapted from Sánchez et al. 2010.

2.3.2. SALL function during wing development

salm is activated by Dpp trough a specific enhancer that is expressed in the central region of the wing including veins LII, LIII and LIV (Barrio & de Celis 2004; De Celis & Barrio 2000). Sall proteins mediate the effect of Dpp signaling during wing development by repressing or activating a large number of target genes (Organista et al., 2015). Removal or exogenous expression of *sall* genes in the wing imaginal discs provoke phenotypes that can be detected in the adult wing (Fig. 9).

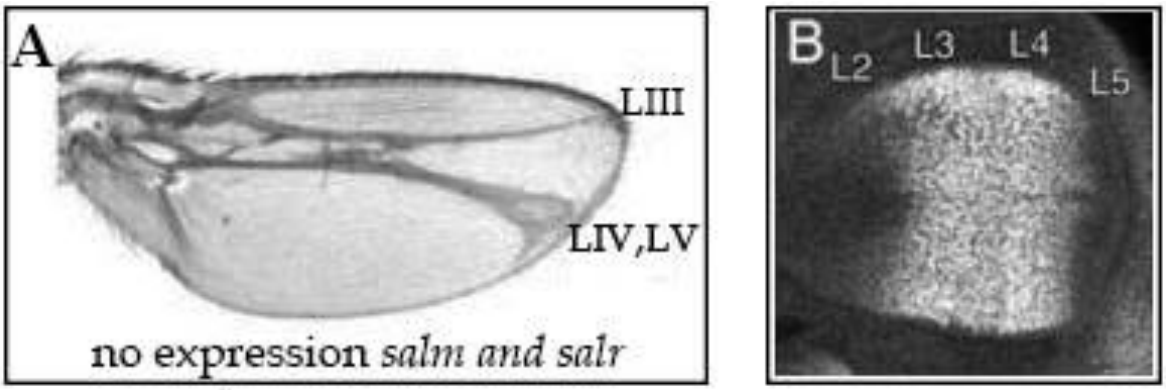


Figure 9. *sall* genes expression contribute to vein patterning. A. When *salm and salr* genes are not expressed vein LII is not formed and veins LIV and LV are fused. **B.** Immunostaining of WT third instar wing imaginal disc. Vein LII forms in the anterior compartment in the region, where Salm is expressed at low levels. The high levels of Salm coincide with the respective area of veins LIII and LIV. In the posterior compartment, Salm expression is adjacent to the corresponding area of vein LV. Adapted from Barrio & de Celis 2004.

SUMOylation of Salm is necessary for promoting ectopic vein LIII in the LII/LIII intervein region when ectopically expressed (Sánchez et al., 2010). It also promotes changes in the localization of Salm in the cells of the wing disc, showing a diffuse distribution when SUMOylation is enhanced (Sánchez et al., 2010). Conversely, Salr shows the opposite behavior, as SUMOylation promotes the formation of large aggregates when expressed in cells and lack of SUMOylation promotes ectopic vein formation when overexpressed in the wing (Sánchez et al., 2010). The stereotypical function of Sall proteins and SUMOylation in the wing can be used as a paradigm to study changes in Sall proteins function/SUMOylation *in vivo*.

II. OBJECTIVES

The long term aim of the laboratory was to identify SUMOylated proteins in steroidogenic tissues. We focused in the following specific objectives:

I. To develop new technology for the isolation and identification of SUMOylated proteins in *Drosophila* cells and *in vivo.*

II. To generate new tools for the analysis of post-translational modifications by UbLs in mammalian cells.

III. To demonstrate that SALL transcription factors are SUMOylated in cells.

IV. To identify possible SUMO E3 ligases for SALL factors, focusing on the Pc family of ligases.

III. HYPOTHESIS

In the first part of this Thesis, I focused on the development of methods to study SUMOylation and other post-translational modifications *in vivo* and in cells. Many strategies in the last 20 years were used in different models to perform global screenings for the identification of SUMOylated proteins. Recently, a new methodology based on the *in vivo* biotinylation of modified proteins was employed in the ubiquitination field (Franco et al., 2011). We hypothesized that the same approach could be used for in the identification of SUMOylated proteins *in vivo*, taking advantage by the strong interaction biotin/streptavidin. The combination of this strategy with Mass Spectrometry would allow us to identify modified proteins in a particular tissue. We also hypothesized that the same methodology could be applied in mammalian cells and for other Ubiquitin-like post-translational modifications.

In the second part of this Thesis, I focused on the SUMOylation of SALL proteins, based on the existing evidences *in vitro* for the *Drosophila* and human homologs (Netzer et al., 2002; Sánchez et al., 2010). We hypothesized that the human and *Drosophila* homologs were SUMOylated in cells and that E3 ligases might be involved in this process. Based on colocalization and interaction analysis by Mass Spectrometry, we hypothesized that the E3 ligases of the Pc family might be involved in SALL1 SUMOylation, Pc in the case of *Drosophila* and CBX4 in humans.

IV. MATERIALS & METHODS

1. Generation of vectors

The following vectors were used in this study (Table 1). Those constructs for this work were done by Dr. J. D. Sutherland, unless otherwise specified.

Name	Reference	Parental vectors	Cloning sites/notes
CAG-bioSUMO1-T2A-BirA^{opt}-T2A-Ubc9	This work	*CAG-bioSUMO1-T2A-BirA^{opt}-T2A-GFPpuro*	GFPpuro exchanged for *UBC9* using *Nhe*II-*Xho*II (AMP)
CAG-bioSUMO3-T2A-BirA^{opt}-T2A-Ubc9	This work	*CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro*	GFPpuro exchanged for *UBC9* using *Nhe*II-*Xho*II (AMP)
CAG-bioUbL-T2A-BirA^{opt}-T2A-GFPpuro (*)	This work	*CAG-bioNEDD8-T2A-BirA^{opt}-T2A-GFPpuro*	*Eco*RI-*Not*I bioUbL; internal *Bam*H1-*Bgl*II for concatemerization (AMP)
CAG-bioUFM1-T2A-BirA^{opt}-T2A-UFC1	This work	*CAG-bioUFM1-T2A-BirA^{opt}-T2A-GFPpuro*	GFPpuro exchanged for UFC1 using *Nhe*II-*Xho*II (AMP)
CAG-BirA^{opt}-T2A-GFPpuro	This work	*CAG-bioNEDD8-T2A-BirA^{opt}-T2A-GFPpuro*	*Acc*65I-*Hind*III; bioNEDD8-T2A removed, Kozak added (AMP)
CAG-BirA^{opt}-T2A-puro	This work	*CAG-BirA^{opt}-T2A-GFPpuro*	GFPpuro exchanged for puro using *Nhe*I-*Xho*I (AMP)
CAG-BirA^{opt}-T2A-Ubc9	This work	*CAG-BirA^{opt}-T2A-GFPpuro*	GFPpuro exchanged for *UBC9* using *Nhe*II-*Xho*II (AMP)
CAG-Myc-BirA-SALL1*	This work	*CAG-Myc-BirA*-2A-GFPpuro*	*Not*I-*Xba*I; (*SALL1* has stop codon) (AMP)
CAG-Myc-BirA-SALL1(826)*	This work	*CAG-Myc-BirA*-2A-GFPpuro*	*Not*I-*Xba*I; (*SALL1* has stop codon) (AMP)
CB6-GFP-SALL1	This work	*CB6-GFP-N*	CB6 has CMV promoter and confers neo selection; contains GFP and MCS; *SALL1* generated by high-fidelity PCR (AMP)

CB6-HA-CBX4	This work	*CB6-HA-N*	*Not*I-*Eco*RI (CB6 has CMV promoter and confers neo selection) (AMP)
CB6-HA-N	M. Way lab (CRUK, London)	*CB6*	CB6 has CMV promoter and confers neo selection; contains HA epitope and MCS (AMP)
*CMV-CBX4-BirA**	This work	*CMV-SALL1-BirA**	*CBX4* (PCR amplified) exchanged for *SALL1* using *Eco*R1-*Sal*1 (KAN)
CMV-CFP-HisMycSalm	Sánchez et al., 2010	*pEGFP-C1*	(KAN)
CMV-CFP-HisMycSalm^{IKDP}	(Sánchez et al., 2010)	*pEGFP-C1*	(KAN)
CMV-EGFP-β-galactosidase	This work	*pEGFP-N1*	LacZ subcloned from pIND/lacZ (Invitrogen)
*CMV-Pc-BirA**	This work	*CMV-SALL1-BirA**	*Drosophila Pc* (PCR amplified) exchanged for SALL1 using *Eco*R1-*Sal*1 (KAN); Pc source: Addgene #1927
CMV-SALL1-2xHA	This work	*CMV-SALL1-YFP*	YFP exchanged for 2xHA using *Sal*I-*Not*I (KAN)
CMV-SALL1-4KR-2xHA	This work	*CMV-SALL1-4KR-YFP*	YFP exchanged for 2xHA using *Sal*I-*Not*I (KAN)
CMV-SALL1-4KR-YFP	This work	*CMV-SALL1-YFP*	*Eco*RI-*Sal*I; mutants introduced by overlap extension PCR (KAN)
*CMV-SALL1-BirA**	This work	*CMV-SALL1-YFP*	YFP exchanged for BirA* using *Sal*1-*Not*1 (KAN)
CMV-SALL1-GFP	(Netzer et al., 2006)	*pEGFP-C1*	(KAN)
CMV-SALL1-YFP	This work	*pEYFP-N1*	*Eco*RI-*Sal*I (KAN); SALL1 generated by high-fidelity PCR
CMV-SALL1(826)-2xHA	This work	*CMV-SALL1(826)-YFP*	YFP exchanged for 2xHA using *Sal*I-*Not*I (KAN)

CMV-SALL1(826)-YFP	This work	*pEYFP-N1*	*Eco*RI-*Sal*I (KAN); SALL1 (C826T) generated by high-fidelity PCR
pAc5-10xUAS-bioSmt3-T2A-BirA^{opt}-T2A-lwr	This work	*pAc5-10xUAS-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro*	GFPpuro exchanged for *lwr* using *Nhe*I-*Xho*I (AMP)
pAc5-10xUAS-FlagCherry-T2A-BirA^{opt}-T2A-lwr	This work	*pAc5-10xUAS-FlagCherry-T2A-BirAopt-T2A-GFPpuro*	GFPpuro exchanged for *lwr* using *Nhe*I-*Xho*I (AMP)
pAc5-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro	This work	*pAc5-STABLE2-neo*	*Acc65I-Not1 (bioSmt3); XbaI-Hind*III (BirA^{opt}); Neo exchanged for GFPpuro using *Nhe*I-*Xho*I (AMP)
pAc5-BirA^{opt}-T2A-GFPpuro	This work	*pAc5-STABLE1-neo*	*Acc*65I-*Hind*III (BirA^{opt}); Neo exchanged for GFPpuro using *Nhe*I-*Xho*I (AMP)
pAc5-Gal4	Addgene #24344	*pAc5c*	(AMP)
pcDNA3	Invitrogen	*pcDNA3*	(AMP)
pEGFP-N1, pEYFP-N1	Clontech	*commercial*	(KAN)
pEYFP-C1, pEGFP-C1	Clontech	*commercial*	(KAN)
pUAST-bioSmt3-T2A-BirA	This work	*pUASTattb*	*Acc*65I-*Hind*III (AMP)
pUAST-bioSmt3::BirA	This work	*pUAST*	This plasmid was constructed by So Young Lee (AMP)
pUAST-BirA	(Franco et al., 2011)	*pUAST*	(AMP)
pUAST-CFP-HisMycSalm	(Sánchez et al., 2010)	*pUAST*	(AMP)

pUASTattB-Flag-βFtz-f1	(Talamillo et al., 2013)	*pUAST-attB*	(AMP)

Table 1. Vectors used in this study.
(*) *CAG-bioSUMO1-T2A-BirA^{opt}-T2A-GFPpuro, CAG-bioSUMO2-T2A-BirA^{opt}-T2A-GFPpuro, CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro, CAG-bioUb-T2A-BirA^{opt}-T2A-GFPpuro, CAG-bioNEDD8-T2A-BirA^{opt}-T2A-GFPpuro, CAG-bioFAT10-T2A-BirA^{opt}-T2A-GFPpuro, CAG-bioISG15-T2A-BirA^{opt}-T2A-GFPpuro* and 2 *CAG-bioUFM1-T2A-BirA^{opt}-T2A-GFPpuro.* MCS: multiple cloning sites; (AMP): ampicillin resistant; (KAN): Kanamycin resistant.

2. Cell culture

Drosophila S2R+ and Kc167 cells were obtained from DGRC (https://dgrc.cgb.indiana.edu) and cultured at 25ºC in an incubator without CO_2 in *Drosophila* Schneider's medium (Invitrogen) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Gibco).

HEK (human embryonic kidney) 293 FT cells (Invitrogen) were maintained at 37ºC with 5% of CO_2 in DMEM (Dulbecco's modified Eagle's medium; Gibco) supplemented with 10% FBS, 1% penicillin/streptomycin (GIBCO).

2.1. Cell transfections

S2R+ Drosophila cells were plated at 50% of confluence. Transient transfections were performed using calcium phosphate in 5 dishes of 10 cm with 10 μg of DNA in complete medium supplemented with 50 μM of biotin using different sets of plasmids:

- 2 μg of *pAc5-Gal4* and 8 μg of *pUAST-bioSmt3::BirA* or *pUAST-BirA.*
- 10 μg of *pAc5-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro* or *pAc5-BirA^{opt}-T2A-GFPpuro.*
- 2 μg of *pAc5-Gal4* and 8 μg of *pAc5-10xUAS-bioSmt3-T2A-BirA^{opt}-T2A-lwr* or *pAc5-10xUAS-FlagCherry-T2A-BirA^{opt}-T2A-lwr.*
- 5 μg of *pUASTattB-Flag-βFtz-f1* and 5 μg of *pAc5-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro* or *pAc5-BirA^{opt}-T2A-GFPpuro.*

HEK 293FT cells were plated at 25-30% of confluence the day before to have 50% confluence at the moment of transfection. Transient transfections were performed using calcium phosphate in 10 cm dishes with 10 μg of DNA in complete medium supplemented with 50 μM of biotin using different sets of plasmids:

- Experiments for isolation of bioSUMO3-conjugates (*Results 2.2., Part I*) coupled with Mass Spectrometry analysis were performed in 10 dishes of 10 cm using 10 μg of *CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro* or *CAG-BirA^{opt}-T2A-GFPpuro.*
- Experiments for isolation of UFM1-conjugates (*Results 2.5, Part I*) coupled with Mass Spectrometry analysis were performed in 10 dishes of 10 cm using 10 μg of *CAG-bioUFM1-T2A-BirA^{opt}-T2A-UFC1* or *CAG-BirA^{opt}-T2A-puro*.
- BioID experiments (*Results 3.1, Part II*) for isolation of possible interactors of SALL1 and SALL1(826) coupled with MS analysis were performed using 10 dishes of 10 cm with 10 μg of *CAG-Myc-BirA*-SALL1* or *CAG-Myc-BirA*-SALL1(826)*.
- For BioID validation (*Results 3.1.1, Part II*) 1 dish of 10 cm was transfected with 10 μg of total DNA mixing 5 μg of *CMV-CBX4-BirA** or *CMV-Pc-BirA** in combination with 5 μg of *CMV-SALL1-2xHA* or *CMV-SALL1(826)-2xHA.* Cells transfected with 5 μg of each plasmid separately were used as controls. The empty vector *pcDNA3* was used as a filler to have cells transfected with same amount of DNA.
- Analysis of SALL SUMOylation (*Results 1.1 and 2.1, Part II*) was executed using cells in a 10 cm dish transfected with the following plasmids: 7 μg of *CMV-SALL1-YFP, CMV-SALL1-2xHA, CMV-SALL1-4KR-2xHA, CMV-CFP-HisMycSalm* or *CMV-CFP-HisMycSalm^{IKDP}* and 3 μg of *CAG-bioSUMO-T2A-BirA^{opt}-T2A-GFPpuro* or *CAG-BirA^{opt}-T2A-GFPpuro*.

For functional analysis of SALL1-CBX4 interaction experiments (*Results 3.3.1, Part II*) 1 dish of 10 cm was transfected with 5 μg of *CMV-SALL1-YFP*, 2.5 μg of *CAG-bioSUMO1-T2A-BirA^{opt}-T2A-GFPpuro* or *CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro* and 2.5 μg of *CB6-HA-CBX4.* The empty vector *pcDNA3* was used as a filler to have cells transfected with same amount of DNA.

2.2. NeutrAvidin pulldown

Transfected cells were collected after 24-48 hours (HEK 293FT) or 3 days (S2R+), washed 3 times with 1x phosphate buffered saline (PBS) and lysed in 1 ml of Lysis Buffer per 10 cm dish [8 M urea, 1% SDS, 50 mM of deSUMOylation inhibitor N-ethylmaleimide in PBS and 1x cOmplete Mini-EDTA free protease inhibitor (Roche)]. Samples were sonicated two times at 10 microns for 15 seconds without overheating the samples. After 20 minutes of top-table centrifugation at maximum speed, samples were diluted by adding 0.5 (HEK 293 FT) or 3 volumes (S2R+) of Binding Buffer [3 M urea, 1 M NaCl, 0.25% SDS] and incubated overnight at room temperature (RT) with 50 µl per dish of a suspension of high capacity NeutrAvidin-agarose resin (ThermoScientific).The following day the supernatant was carefully separated from the beads and saved as the flow-through fraction (FT).Washes were done with different Washing Buffers (WB) using 10 times the beads' volume, according to Franco and coworkers (Franco et al., 2011): 2x WB1, 3x WB2, 1x WB3, 3x WB4, 1x WB1, 1x WB5 and 3x WB6. Tubes were inverted 3-4 times, centrifuged at 1000 rpm and the supernatants were discarded with gentle aspiration after every wash. WB1: 8 M urea, 0.25% SDS in PBS; WB2: 6 M guanidine hydrochloride in PBS; WB3: 6.4 M urea, 1 M NaCl, 0.2% SDS in PBS (pre-warmed); WB4: 4 M urea, 1 M NaCl, 10% isopropanol, 10% ethanol, 0.2 % SDS in PBS; WB5: 8 M urea, 1% SDS in PBS; WB6: 2% SDS in PBS.

For the elution, samples were heated for 5 minutes at 95ºC in 4x Laemmli sample buffer with 100 mM DTT. The eluted samples were separated carefully from the beads after 5 minutes of centrifugation at maximum speed.

2.3. GFP-Trap co-pulldown

HEK 293FT cells were plated at 25-30% of confluence. Transient transfections were performed using calcium phosphate in a 10 cm dish with 5 µg of *CMV-SALL1-GFP* and 5 µg of *CB6-HA-CBX4* or *CB6-HA* in complete medium. Differently from NeutrAvidin pulldowns, the protocol for GFP-Trap should be performed at 4°C. Two days after transfection, cells were washed 2 times with cold 1x PBS and detached from the dish with a scraper. Cells of 10 cm dishes were lysed by adding 1 ml of Lysis Buffer followed by incubation on a rotating wheel for 30 minutes at 4°C. After 15 minutes of top-table full speed centrifugation at 4°C, supernatants were transferred to a new tube and incubated on a rotating wheel overnight at 4°C with 25 µl of suspension beads (GFP-Trap, ChromoTek) previously washed 2 times with Lysis Buffer. After incubation, beads were washed 5 times with 500 µl of WB. For the elution, 35 µl of 2x Laemmli

buffer were added to beads. After 5 minutes at 95°C and centrifugation at top speed, samples were analyzed by Western blot. Lysis Buffer: 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% TritonX-100, 5% glycerol and 1x cOmplete Mini-EDTA free protease inhibitor (Roche). WB: 25 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% TritonX-100 and 5% glycerol.

2.4. Immunofluorescence in cultured cells

For immunostaining and microscopy, cells were placed on 12 mm diameter round coverslips that were prepared by immersion in a solution of 1 N HCl overnight at RT. After several washes with H_2O to remove the HCl, coverslips were dried and sterilized.

S2R+ *Drosophila* cells are considered semi-adherent and, to prevent their detachment during the immunofluorescence process, we used coverslips treated with ConcanavalinA (ConA, Sigma). We incubated the coverslips with a sterile solution of 0.5 mg/ml of ConA for 15 min. After that, we washed them once with 1x PBS and we placed them in 24-well plates. We transfected the S2R+ cells in 6-well plates with 2 μg of total DNA using calcium phosphate with the following plasmids: 2 μg of *pAc5-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro* or *pAc5-Gal4* and 1 μg of *pUAST-CFP-HisMycSalm* with 1 μg of *pAc-Gal4.* After 3 days, we transferred the transfected cells to 24-well plates with ConA treated-coverslips and let them attach overnight to be ready for processing.

U2OS adherent mammalian cells were plated directly in 24-well plates with coverslips HCl-treated and transfected using Linear PEI (MW 25 kDa; Polyscience #23966) according to the manufacturer's instructions with different set of plasmids:

- Experiments for the localization of different bioUbLs (*Results 2.3, Part I*): 2 μg of *CAG-bioUbL-T2A-BirA^{opt}-T2A-GFPpuro* or 2 μg of *CAG-BirA^{opt}-T2A-puro* as a control.
- Experiments for SUMOylation of SALL proteins (*Results 1.2, 1.2.1, 2.2 and 3.2.1, Part II*): 2 μg of *CB6-GFP-SALL1* or *pEGFP-N1*, 2 μg of *CMV-SALL1-YFP* or *pEYFP-C1*, 2 μg of *CMV-SALL1-4KR-YFP* or *pEYFP-C*, 1.5 μg of *CMV-SALL1-YFP* with 0.5 μg *CAG-bioSUMO1-T2A-BirA^{opt}-T2A-Ubc9, CAG-bioSUMO3-T2A-BirA^{opt}-T2A-Ubc9* or *CAG-BirA^{opt}-T2A-Ubc9*, 2 μg *CMV-CFP-HisMycSalm* or *CMV-CFP-HisMycSalm^{IKDP}*.

After 2 days, cells transfected were ready to be processed for immunofluorescence. Wells with transfected cells (either S2R+ or U2OS) were washed 3 times with 1x PBS to remove

dead cells. Cells were fixed in 4% paraformaldehyde (Santa Cruz) for 20 min, after which they were washed twice with 1x PBS. Permeabilization was performed with 0.1% Triton X-100 in 1x PBS for 15 min. Coverslips were then washed 3 times with 1x PBS to remove the detergent. Blocking was performed using 2% bovine serum albumin (BSA) and 2% fetal bovine serum (Gibco) in 1x PBS for 1 h at RT. Incubation with primary antibodies diluted in blocking solution was performed during 1 h at 37°C in a humidity chamber or overnight at 4°C. The following primary antibodies were used in S2R+ *Drosophila* cells: mouse monoclonal anti-GFP (1:500, Roche #11814460001); rabbit polyclonal anti-Smt3 (1:150; Smith et al. 2004); rabbit polyclonal anti-Pc (1:200; Gonzalez et al. 2014). The following primary antibodies were used in U2OS cells: rabbit polyclonal anti-SALL1 (1:200, Abcam #31905); mouse monoclonal anti-GFP (1:500, Roche #11814460001); mouse monoclonal anti-PML (1:100, Santacruz #sc-966); mouse monoclonal anti SC-35 (1:100, BD biosciences #556363); rabbit polyclonal anti-CBX4 (1:100, Proteintech #18544-1-AP); rabbit polyclonal anti-SUMO2/3 (1:100, Eurogentec #AV-SM23-0100); mouse monoclonal anti SUMO1 (1:100, The Developmental Studies Hybridoma Bank, DSHB, #21C7); mouse monoclonal anti-SUMO2 (1:100, DSHB #8A2).

After incubation with the primary antibody, cells were gently washed 3 times with 1x PBS and then incubated with the secondary antibody in the dark for 1 h at RT. The secondary antibodies conjugated to fluorofore used were: anti-mouse or anti-rabbit Alexa Fluor 488, Alexa Fluor 568 or Alexa Fluor 647 (1:200, Molecular Probes); Alexa Fluor 594 streptavidin-conjugated (1:200, Jackson ImmunoResearch). To visualize the nuclei we incubated the cells with DAPI (1:15000, Roche #10236276001) for 5 minutes at RT. Another 3 washes were performed to remove unbound secondary antibody. Finally, coverslips were mounted using Prolong Gold antifade reagent (Molecular Probes #P36930) and stored in the dark at 4°C.

Stained cells were visualized using an Upright Fluorescent Microscope Axioimager D1 or a Leica DM IRE2 confocal microscope using Blue diode, Blue Ar/ArKr, DPSS and HeNe lasers for excitation, and 63x lens for magnification. Pictures were analyzed with the Leica confocal software and the Adobe Photoshop program.

For functional analysis of SALL1-CBX4 interaction (*Results 3.3.2, Part II*), U2OS cells were transfected in 24 well plate with 500 ng of *CMV-SALL1-YFP*, 500 ng *CMV-SALL1(826)-YFP* or 500 ng *CMV-EGFP-β-galactosidase.* After 2 days, cells were processed for immunofluorescence as explained above but directly in the well and analyzed with an automated fluorescent microscope ImageXpress (Molecular Devices). Rabbit polyclonal anti-CBX4 (1:100, Proteintech #18544-1-AP), Alexa Fluor 568 (1:200, Molecular Probes) and DAPI (1:15000, Roche #10236276001) were used. 64 pictures per well were taken to detect SALL proteins or β-

galactosidase as control (green), CBX4 (red) and DAPI (blue). The analysis was performed using the Multi-Wavelenght-Cell-Scoring Application Module of MetaXpress image analysis software (Molecular Devices). PcG (Protein group) bodies were counted in transfected cells. These experiments were performed in triplicates. The statistical analysis was done taking into account a minimum of 100 cells per condition. One Way ANOVA analysis was performed using GraphPad Prism 5 software.

3. Western blot

Input, FT and Elution samples were generated at different steps of the pulldowns. As described before, samples were eluted by adding Laemmli buffer 4x supplemented with 100 mM of DTT directly to the beads. Input and FT samples were prepared by mixing with the same buffer. Samples were boiled at 95° for 5 minutes before being loaded into the gel. Proteins were separated on SDS polyacrylamide gels using a Mini-PROTEAN system (BioRad). The gels were prepared as shown in the following tables, adjusting for gel percentage according to protein target size.

GEL%	7.5%	10%	12.5%	STACK 4%
30% Acrylamide/Bis (BioRad #1610158)	1.352 ml	1.81 ml	2.25 ml	0.54 ml
0.5 M Tris-HCl pH 6.8				0.75 ml
1.5 M Tris-HCl pH 8.8	1.325 ml	1.325 ml	1.325 ml	
10% SDS (Fluka # 05030)	55 µl	55 µl	55 µl	30 µl
TEMED (Sigma #GE17-1312-01)	6 µl	6 µl	6 µl	3 µl
10% Ammonium persulfate (Sigma #A3678)	27.5 µl	27.5 µl	27.5 µl	15 µl
H_2O	2.4 ml	2.23 ml	1.750 ml	1.650 ml

Table 2. Composition of polyacrilamide gels.

For isolation of UbL-conjugates, Mini-PROTEAN precast gradient gels at 4-15% were used (BioRad). Gels were run at 80-100 V for 1 h and transferred to PVDF membranes

(Millipore) at 70-80 V/gel for 1,5 h using the wet system of BioRad. Membranes were blocked in 1x PBS with 0.1% Tween-20 (PBS-T) and 5% non-fat dry milk (blocking buffer) for 1 hour, or with Casein Blocking Buffer 1x (Sigma #B6429), necessary when the anti-biotin antibody was used. After that, membranes were incubated in blocking buffer for 1 hour at RT or overnight at 4°C with the following primary antibodies: chicken polyclonal anti-BirA (1:1000, Abcam #ab14002); mouse monoclonal anti-Flag M2 (1:1000, Sigma #F3165); mouse monoclonal anti-USP (1:200; Christianson et al. 1992); mouse monoclonal anti-Osa (1:50, DSHB); mouse monoclonal anti-Lamin Dm0 (1:500, DSHB # ADL84.12); rabbit polyclonal anti-Fax (1:1000; Gates et al. 2009); rabbit polyclonal anti-eIF4E (1:1000; Lachance et al. 2002); rabbit polyclonal anti-SIRT1 (1:1000, Cell Signaling Technology #D739); mouse monoclonal anti-RanGAP1C-5 (1:1000, SantaCruz #sc-28322); rabbit polyclonal anti-PARP (1:1000, Cell Signaling Technology #9542); rabbit polyclonal anti-PML (1:1000, Bethyl Laboratories); rabbit polyclonal anti-Ub (1:1000, Sigma #U5379); mouse monoclonal anti-GFP (1:1000, Roche #11814460001); mouse monoclonal anti-HA (1:1000, Sigma #H3663); mouse monoclonal anti-actin AC-74 (1:5000, Sigma #A2228).

After three washes with PBS-T, the blots were incubated for 1 h in blocking buffer with secondary antibodies: HRP-conjugated anti-mouse (1:5000, Jackson ImmunoResearch); HRP-conjugated anti-rabbit (1:5000, Jackson ImmunoResearch); HRP-conjugated anti-chicken (1:2000, Abcam #ab97135); HRP conjugated anti-biotin (1:1000, Cell Signaling Technology #7075); HRP conjugated anti-tubulin (1:5000, Proteintech #66031); HRP conjugated anti-GAPDH (1:5000, Proteintech #60004). Membranes were washed again three times in PBS-T and then developed using chemiluminescence with ECL Western Blotting Detection Reagents (Amersham GE Healthcare) or Clarity Western ECl substrate (Bio-Rad).

4. In vitro SUMOylation and SUBEs pulldown

In vitro SUMOylation assays were performed by Dr. Valerie Lang in the Laboratory of Dr. Manuel S. Rodriguez (Inbiomed, San Sebastian). SALL1-2xHA and CBX4 were transcribed/translated *in vitro* using the TNT ® Quick Coupled Transcription/Translation System (Promega) according with the manufacturer's instruction and then were incubated in a buffer containing an ATP regenerating system [(50 mM Tris pH 7.5, 10 mM MgCl2, 2 mM ATP, 10 mM creatine phosphate (Sigma), 3.5 U/ml of creatine kinase (Sigma), and 0.6 U/ml of inorganic

pyrophosphatase (Sigma)], 10 µg of SUMO1 or a combination of 5 µg of SUMO2 and SUMO3, 0.325 µg Ubc9 and 0.8 µg of purified SAE1/2 (ENZO Life Sciences). SALL1 or CBX4 SUMOylation were checked adding 2 µl of *in vitro* transcribed/translated protein in the SUMOylation assay. Reactions were incubated at 30°C for 2 hours and stopped by addition of SDS sample buffer. For pulldown assays, 1/10 of the input was saved and the rest of the reaction was incubated with 75 µl of GST-agarose beads containing 75 µg of SUBEs (Da Silva-Ferrada et al., 2013) and 1 mM DTT for 2 h, at 4°C. After incubation, beads were pulled down by centrifugation for two minutes at 1000 rpm at 4°C. Subsequently, the beads were washed with 30 column volumes of binding buffer (50 mM Tris pH 8.5, 50 mM NaCl, 5 mM EDTA and 1% Igepal) and were resuspended in one column volume of 2x Laemmli buffer.

5. Proximity Ligation Assays

The experiments of this part of the PhD Thesis were done during a short term stay in the laboratory of Dr. A. Vertegaal (LUMC, Leiden, Netherlands) and were supported by a grant from the COST Action PROTEOSTASIS (BM1307). U2Os cells were plated and transfected by P-PEI in 6-well plates with 2 µg of *CMV-SALL1-2xHA, CMV-SALL1-4KR-2xHA or pcDNA3.* After 2 days, cells were transferred to a 8-well chamber slide (LabTek #177410) and let them attach for 12 hours. Proximity Ligation Assay (PLA) was performed using the Duolink In Situ Red kit (Olink Biosciences; Gullberg et al. 2004; Söderberg et al. 2006) according to the manufacturer's instructions. Primary antibodies used: mouse monoclonal anti-SALL1 (1:250, R&D Systems #K9814); rabbit polyclonal anti-SUMO2/3 (1:250; Vertegaal et al. 2004); rabbit polyclonal anti CBX4 (1:100, Proteintech #18544-1-AP). Images were recorded on a Leica SP8 confocal microscope system using 488 nm and 561 nm wavelengths for excitation and a 63x lens for magnification, and were analyzed with the Leica confocal software, Adobe Photoshop and ImageJ software.

6. Mass Spectrometry

Mass Spectrometry experiments were realized by Jón Otti Sigursson in the laboratory of Dr. Jesper Olsen (University of Copenhagen, Denmark). Pulldown elutions were separated in SDS-page and stained with Brilliant Blue G-Colloidal Concentrate (Sigma #B2025) according

with manufacturer's instructions. Proteins were in-gel digested, extracted, concentrated and analyzed using and EASY nLC system (Proxeon) connected to Q exactive orbitrap (ThermoFisher) trough a nanoelectrospray ion source. Raw Mass Spectrometry files were processed with a MaxQuant software (version 1.4.0.3, Max Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction).

The lists of proteins identified by Mass Spectrometry were analyzed as follows. First, contaminants and proteins identified by only one peptide were eliminated. Then, only those proteins with 4 fold of iBAQ (intensity-based absolute quantification) in the experiment samples *versus* the controls were considered as positive hits. Calculation was done taking into account a baseline for the control, which corresponds to the minimum value of iBAQ registered in each control of every experiment.

The functional interpretation of the data was performed with Vlad free access software for *Drosophila* proteins (proto.informatics.jax.org/prototypes/vlad-1.0.3) and with innateDB for the mammalian proteins (Breuer et al., 2013).

7. *Drosophila* husbandry and stocks

Fly stock were raised on standard *Drosophila* medium (flour, yeast, glucose, polenta, agar and propionic acid) at 25°C. Genetic crosses were performed generally at 25°C, except when indicated at 18°C or 29°C.

Binary system Gal4/UAS (Brand and Perrimon, 1993) was used to direct tissue-specific gene expression. The following Gal4 lines were used:

Genotype	Name used in the text	Reference
w;phm-Gal4,UAS-mCD8::GFP/TM6B,Tb	*phm-Gal4*	Mirth et al. 2005
w^{*}; P{Gal4-Hsp70.PB}89-2-1	*hs-Gal4*	Bloomington #1799
w^{*};P{Gal4-ninaE.GMR}12	*GMR-Gal4*	Bloomington #1104
w;sal^{EPV}-Gal4	*Sal^{EPV}-Gal4*	Barrio & de Celis 2004

Table 3. Gal4 lines used in this work

phm-Gal4 drives expression in the prothoracic gland (Mirth et al., 2005); *GMR-Gal4* drives expression in the developing eye; *hs-Gal4* drives expression in the whole body after heat shock (hs) at 29°C or 37°C; *Sal^{EPV}-Gal4* drives expression in the wing disc regulated by a *salm* wing-specific enhancer (Barrio and de Celis, 2004).

Table 4 shows the *UAS* lines used for RNA interference or overexpression.

Genotype	Name used in the text	Reference
w;If/CyO-GFP;UAS-bioSmt3::BirA/TM6B	*UAS-bioSmt3::BirA*	This work
w;UAS-smt3i8/CyO-GFP;UAS-biosmt3::BirA,phm-Gal4/TM6B	*w;UAS-Smt3i,UAS-bioSmt3::BirA/CyO-GFP;phm-Gal4/TM6B*	This work
w,GMR-Gal4,UAS-smt3i8/CyO;TM2/TM6B	*GMR-Gal4,UAS-smt3i8/CyO;TM2/TM6B*	Ugo Mayor
w;Bl/CyO-GFP-GFP;UAS-BirA/TM6B	*w;Bl/Cy-GFPO;UAS-BirA/TM6B*	Ugo Mayor
w;UAS-Smt3i/CyO-GFP;UAS-bioSmt3-T2A-BirA^{opt}/TM6B	*UAS-Smt3i/CyO;UAS-bioSmt3-T2A-BirA^{opt}/TM6B*	This work
w;UAS-Smt3i/CyO-GFP;UAS-BirA	*w;UAS-Smt3i/CyO-GFP;UAS-BirA*	This work
w;UAS-Pc-RNAi/TM6B	*UAS-Pci*	NIG-FLY #32443R-1
w;UAS-Ubxi (37825)	*UAS-Ubxi*	VDRC #37825
w;UAS-Ubxi (37823)	*UAS-Ubxi*	VDRC #37823
w;UAS-CFP-sal-6.1/CyO;UAS-nls-eGFP/TM6B	*UAS-salm,UAS-GFP*	Sánchez et al. 2010
w;UAS-CFP-sal-6.1/CyO;smt3i-13/TM6B	*UAS-salm,UAS-smt3i*	Sánchez et al. 2010
w;UAS-nls-eGFP/CyO;UAS-YFP-salR1/TM6B	*UAS-GFP,UAS-salr*	Sánchez et al. 2010
w;UAS-nls-eGFP/CyO;MKRS/TM6B	*UAS-GFP*	Sánchez et al. 2010

w,UAS-CFPsal6.1::smt3deg/CyO;UASnlsGFP/TM6B	*UAS-smt3::salm,UAS-GFP*	Sánchez et al. 2010
w;UAS-CFPsal6.1(IKPD)/CyO; UASnlsGFP/TM6B	*UAS-salm^{IKPD},UAS-GFP*	Sánchez et al. 2010
w;UASnlsGFP/CyO;UAS-YFPsalR1::smt3deg/TM6B	*UAS-GFP,UAS-smt3::salr*	Sánchez et al. 2010
w;UASnlsGFP/CyO;UAS-YFP-salR1^{IKEA,IKVA}/TM6B	*UAS-GFP,UAS-salr^{IKEA,IKVA}*	Sánchez et al. 2010

Table 4. UAS lines used in this work.
NIG-FLY, Fly Stocks of National Institute of Genetics, Japan; VDRC, Vienna *Drosophila* Resource Center.

Table 5 shows the mutant lines used to study the relationship between *salm*, *Pc* and the SUMOylation machinery.

Genotype	Name used in the text	Reference
Pc[3]/TM1	*Pc[3]*	Consolider collection
P{PZ}lwr[05486],cn1/CyO;ry506	*lwr[05486]*	Bloomington #11410
y,w;lwr[4-3],P{ry[+t7.2]=neoFRT}40A/CyO,y+	*lwr[4-3]*	Bloomington #9321
y,w;lwr[5]b[1]cn[1]bw[1]/CyO, y+	*lwr[5]*	Bloomington #9317
P{PZ}smt[304493],cn1/CyO;ry506	*smt3[04493]*	Bloomington #11378
w;Df32FP-5/CyO,w+,GFP	*Df5(2.42)*	Barrio et al. 1999
w;Df32FP-5/CyO,twist-Gal4,UAS-GFP	*Df5(3.04)*	This work

Table 5. *Drosophila* mutant lines used in this work.

7.1. Generation of transgenic flies

To generate *UAS-bioSmt3::BirA* and *UAS-bioSmt3-T2A-BirA* transgenic lines, *pUASTattB-bioSmt3::BirA* and *pUASTattB-bioSmt3-T2A-BirA* vectors (described in *Section 1* of *Materials and Methods*) were used respectively. Injection in embryos and transgenic fly selection were committed to the Consolider *Drosophila* transgenesis service (Genshape project, Consolider CSD 2007-008-25120, CBMSO, Madrid). Transgenic lines were selected, balanced and crossed in collaboration with Dr. Coralia Perez. Adult flies were crossed with *white* mutant flies and flies with red eyes were selected. Those animals were crossed with *w-;Bl/CyO-GFP;TM2/TM6B* balancer line to map the insertion of the transgene and to generate balanced stable lines.

7.2. Isolation of bioSmt3 conjugates *in vivo*

Isolation of SUMOylated proteins from the ring gland was performed using the recombinant line *w;UAS-Smt3i,UAS-bioSmt3::BirA/CyO-GFP;phm-Gal4/TM6B*. Approximately 100 larvae at the L3 instar were dissected in cold 1x PBS. Ring glands attached to brain complexes were selected and lysed in 25 µl of cold Protein extraction Buffer [50 mM of NEM, 1X cOmplete Mini-EDTA free protease inhibitor (Roche) in 1x PBS]. Tissue lysates were used to perform NeutrAvidin pulldowns as described previously using 100 µl of suspension beads. As control, the *w;phm-Gal4/TM6B* line was used.

For the isolation of SUMOylated proteins from adult eyes, we used the recombinant line *w,GMR-Gal4,UAS-smt3i/CyO-GFP;UAS-bioSmt3::BirA/TM6B,* obtained by crossing *w,GMR-Gal4,UAS-smt3i/CyO;TM2/TM6B* and *w;If/CyO-GFP;UAS-bioSmt3::BirA/TM6B*. 20 ml of 10 days-old flies were collected and, with a sieve, heads were separated from the bodies. All passages were done at very cold temperature to avoid protein degradation. Heads were crushed by using an homogenizer and lysed in 2.5 ml of Lysis Buffer (*Section 2.2 of Materials and Methods*). After 2x 5 minutes centrifugation in a top-table centrifuge at maximum speed at 4°C, supernatants were used to perform NeutrAvidin pulldowns by using 200 µl of suspension beads. As control, we used the *w;GMR-Gal4,UAS-smt3i/CyO-GFP;UAS-BirA/TM6B* line obtained from crossing *w;Bl/CyO-GFP;UAS-BirA/TM6B* and *w;GMR,UAS-Smt3i/CyO;TM2/TM6B* lines.

For the isolation of SUMOylated proteins from the whole larva, *w;UAS-Smt3i/CyO;UAS-bioSmt3-T2A-BirA^{opt}/TM6B* line was crossed with *w;hs-Gal4/TM6B* and *w;UAS-Smt3i/+;bioSmt3-T2A-BirA^{opt}/hs-Gal4* larvae were selected at L2 instar, choosing larvae

without GFP and TM6B markers. As controls, we used *w;UAS-Smt3i/+;UAS-BirA/hs-Gal4* larvae selected as explained before from the cross between *w;UAS-Smt3i/CyO-GFP;UAS-BirA* and *w;hs-Gal4/TM6B,* taken at the same larval stage. 50 μM biotin was added to the food. Approximately 200 larvae were collected per condition and three hs events were performed in the following way: first hs at 37°C for 30 minutes at L2 instar stage followed by 12 hours at 25°C, a second hs at 37°C for 30 minutes followed by 6 hours at 25°C, and a third and last hs at 37°C for 30 minutes followed by recovery at 25°C for 2 hours before dissection. To reduce the proportion of non-proliferating tissues during the dissection, only the anterior part of the larvae was collected, where brain, ring gland and most of the imaginal discs are located. Dissections were performed in cold 1x PBS and the lysis was done by using 1 ml of Lysis Buffer (*Section 2.2 of Materials and Methods*). With the help of an homogenizer, larval tissues were smashed and after top speed centrifugation at 4°C, the supernatants were used for NeutrAvidin pulldowns (previously described in *Section 2.2* of *Materials and Methods*) using 100 μl of suspension beads.

7.3. Analysis of adult wing phenotype

To study genetic interaction between *sall*, *Pc* and *smt3*, we collected wings of the following genotypes obtained by crossing the pertinent stocks indicated in Table 6.

Genotype	Name used in the text	Reference
w;Sal^{EPV}-Gal4/+;UAS-Pci/UAS-GFP	*Sal^{EPV}-Gal4>UAS-Pci,UAS-GFP*	This work
w;Sal^{EPV}-Gal4/+;UAS-Pci/UAS-Ubxi	*Sal^{EPV}-Gal4>UAS-Pci,UAS-Ubxi*	This work
w;Sal^{EPV}-Gal4/UAS-salm;UAS-GFP	*Sal^{EPV}-Gal4>UAS-salm,UAS-GFP*	Sánchez et al. 2010
w;Sal^{EPV}-Gal4/UAS-salm;UAS-Pci/UAS-Ubxi	*Sal^{EPV}-Gal4>UAS-salm,UAS-Pci,UAS-Ubxi*	This work
w;Sal^{EPV}-Gal4/UAS-GFP;UAS-salr/+	*Sal^{EPV}-Gal4>UAS-GFP,UAS-salr*	Sánchez et al. 2010
w;Sal^{EPV}-Gal4/UAS-GFP;UAS-	*Sal^{EPV}-Gal4>UAS-GFP,UAS-*	This work

salr/UAS-Pci	*salr,UAS-Pci*	
w;Sal^{EPV}-Gal4/UAS-smt3-salm;UAS-GFP/UAS-GFP+	*Sal^{EPV}-Gal4>UAS-smt3-salm,UAS-GFP,UAS-GFP*	Sánchez et al. 2010
w;Sal^{EPV}-Gal4/UAS-smt3-salm;UAS-Pci/UAS-Ubxi	*Sal^{EPV}-Gal4>UAS-smt3-salm,UAS-Pci,UAS-Ubxi*	This work
w;Sal^{EPV}-Gal4/UAS-salm^{IKPD};UAS-GFP/UAS-GFP	*Sal^{EPV}-Gal4>UAS-salm^{IKPD},UAS-GFP,UAS-GFP*	Sánchez et al. 2010
w;Sal^{EPV}-Gal4/UAS-salm^{IKPD};UAS-Pci/UAS-Ubxi	*Sal^{EPV}-Gal4>UAS-salm^{IKPD},UAS-Pci,UAS-Ubxi*	This work
w;Sal^{EPV}-Gal4/UAS-GFP;UAS-salr/+	*Sal^{EPV}-Gal4>UAS-GFP,UAS-smt3-salr*	Sánchez et al. 2010
Sal^{EPV}-Gal4/UAS-salr;UAS-Pci/+	*Sal^{EPV}-Gal4>UAS-salr,UAS-Pci*	This work
w;Sal^{EPV}-Gal4/UAS-GFP;UAS-smt3-salr/+	*Sal^{EPV}-Gal4>UAS-GFP,UAS-smt3-salr*	Sánchez et al. 2010
w;Sal^{EPV}-Gal4/+;UAS-smt3-salr/UAS-Pci/+	*Sal^{EPV}-Gal4>UAS-smt3-salr,UAS-Pci*	This work
w;Sal^{EPV}-Gal4/UAS-GFP;UAS-salr^{IKEA,IKVA}/+;	*Sal^{EPV}-Gal4>UAS-GFP,UAS-salr^{IKEA,IKVA}*	Sánchez et al. 2010
w;Sal^{EPV}-Gal4/+;UAS-salr^{IKEA,IKVA}/UAS-Pci/+;	*Sal^{EPV}-Gal4>UAS-salr^{IKEA,IKVA},UAS-Pci*	This work
w;Sal^{EPV}-Gal4/UAS-salm; UAS-smt3i;UAS-GFP	*Sal^{EPV}-Gal4>UAS-salm,UAS-smt3i,UAS-GFP*	J. Sánchez, PhD Thesis.
w;Sal^{EPV}-Gal4/UAS-GFP;UAS-Pci/UAS-smt3i	*Sal^{EPV}-Gal4>UAS-GFP,UAS-Pci,UAS-smt3i*	This work
w;Sal^{EPV}-Gal4/UAS-salm;UAS-Pci/UAS-smt3i	*Sal^{EPV}-Gal4>UAS-salm,UAS-Pci,UAS-smt3i*	This work

Table 6. Genotypes used to study interaction between *salm*, *Pc* and *smt3*.

Adult wings were also collected from transheterozygous lines: *Pc[3];lwr[05486]*, *Pc[3];lwr[4-3]*, *Pc[3];lwr[5]*, *Pc[3],smt3[04493]*, *Df-5;Pc[3]*.

Wings were dissected in ethanol at 95% and mounted in a solution of lactic acid:ethanol 50:50 with the help of Dr. Coralia Perez. Pictures were taken with a Canon

PowerShot S70 camera coupled to a Leica DM 1000 microscope with a 4x objective. For size quantification, slides with the mounted wings were scanned in a EPSON scanner at 4800 dpi resolution. Images were treated in Adobe Photoshop and analyzed using ImageJ. For the calculation of the wing size, the squared pixels obtained from ImageJ were multiplied by 4800 x 4800 obtaining the size in µm.

7.4. Immunohistochemistry

Wing imaginal discs and ring glands of L3 larvae were dissected in cold 1x PBS and fixed in 4% paraformaldehyde for 20 minutes, after which they were washed 3 times with PBT (1x PBS, 0.03% Triton X-100) for 20 minutes. Blocking was performed for 1 hour in PBT with 1% of BSA. Incubation with the primary antibodies diluted in blocking solution was done overnight at 4°C. The following antibodies were used: polyclonal anti-Salm (1:200; Barrio et al. 1999); mouse anti-Ubx (1:10; White & Wilcox 1984); rabbit polyclonal anti-Pc (1:200; Gonzalez et al. 2014). The following day, samples were rinsed 3 times in PBT and, after two washes of 30 minutes in PBT, were incubated with the secondary antibodies conjugated to fluorophores for 2 hours in the dark at RT. The following secondary antibodies were used: anti-mouse, anti-rabbit or anti-rat Alexa Fluor 488, Alexa Fluor 568 or Alexa Fluor 647 (1:200, Molecular Probes). Larvae were rinsed again 3 times with PBT and washed 3 times for 20 minutes with PBT. After that, samples were rinsed in 1x PBS and the tissues of interest were dissected, mounted in Vectashield medium (Roche) and visualized by a confocal Leica DM IRE2 microscope. Pictures were processed with the Adobe Photoshop program.

8. Statistical analysis

Statistical analyses were performed with GraphPad Prism 5 software. Student's *t* test or One Way ANOVA were applied to determine significance, as required.

V. RESULTS

PART I: ANALYSIS OF POST-TRANSLATIONAL MODIFICATIONS BY UBIQUITIN-LIKES

1. Identification of SUMOylated proteins in *Drosophila melanogaster*.

Smt3 is necessary for *Drosophila* development and, as mentioned in the *Introduction*, SUMOylation has an important role in *Drosophila* steroidogenesis. Conditional knockdown of *smt3* in the PG stops the larval development at the third instar (Talamillo et al. 2008). Since our long term aim is to identify SUMOylated proteins responsible of this strong phenotype, we developed a system based on the *in vivo* biotinylation of Smt3 for the identification of the Smt3-subproteome.

1.1. New methodology for the isolation of SUMOylated proteins in *Drosophila* cells

The new technology was based on the method implemented by Franco and coworkers (Franco et al., 2011) to identify ubiquitinated proteins *in vivo* in *Drosophila melanogaster* and was developed in collaboration with Dr. J. D. Sutherland in the laboratory. We cloned a specific sequence (*bio*) at the N terminus of *deg-smt3* [a degenerated sequence of *smt3* at the nucleotide level that was insensitive to the dsRNA against *smt3* and, therefore, was still translated when the endogenous *smt3* was silenced (Sánchez et al., 2010)], generating the transgene *bioSmt3*. The *bio* sequence is recognized by BirA, an enzyme from *E. coli* that attaches a biotin molecule to it. We expressed *BirA* and *bioSmt3* together with or without a selectable marker. The bioSmt3 peptide will carry the biotin moiety, rendering SUMOylated targets that are at the same time biotinylated and that will be isolated by NeutrAvidin chromatography (Fig. 10). The very high affinity between biotin and NeutrAvidin allows perform pulldown analysis under stringent conditions, which generates high yields with consequent low background in the experiments.

To check whether this technology was useful in the investigation of SUMOylation in *Drosophila*, first we explored its possibilities in cultured cells. We generated the vector *pUAST-bioSmt3::BirA*, which could be used for the generation of transgenic flies or for the expression in cultured cells in combination with *pAc5-Gal4*. The usage of this vector depended on the correct

release of BirA from bioSmt3 by endogenous Ulp proteases present in the cellular environment after translation of the linear polypeptide. For the NeutrAvidin pulldown, we introduced some important modifications to the protocol developed by Franco and coworkers (Franco et al., 2011). These modifications were required in order to adapt the method to the isolation of SUMOylated proteins, which are less abundant than ubiquitinated proteins and normally represent a small percentage of the total pool of proteins. These modifications are described in detail in *Materials and Methods.*

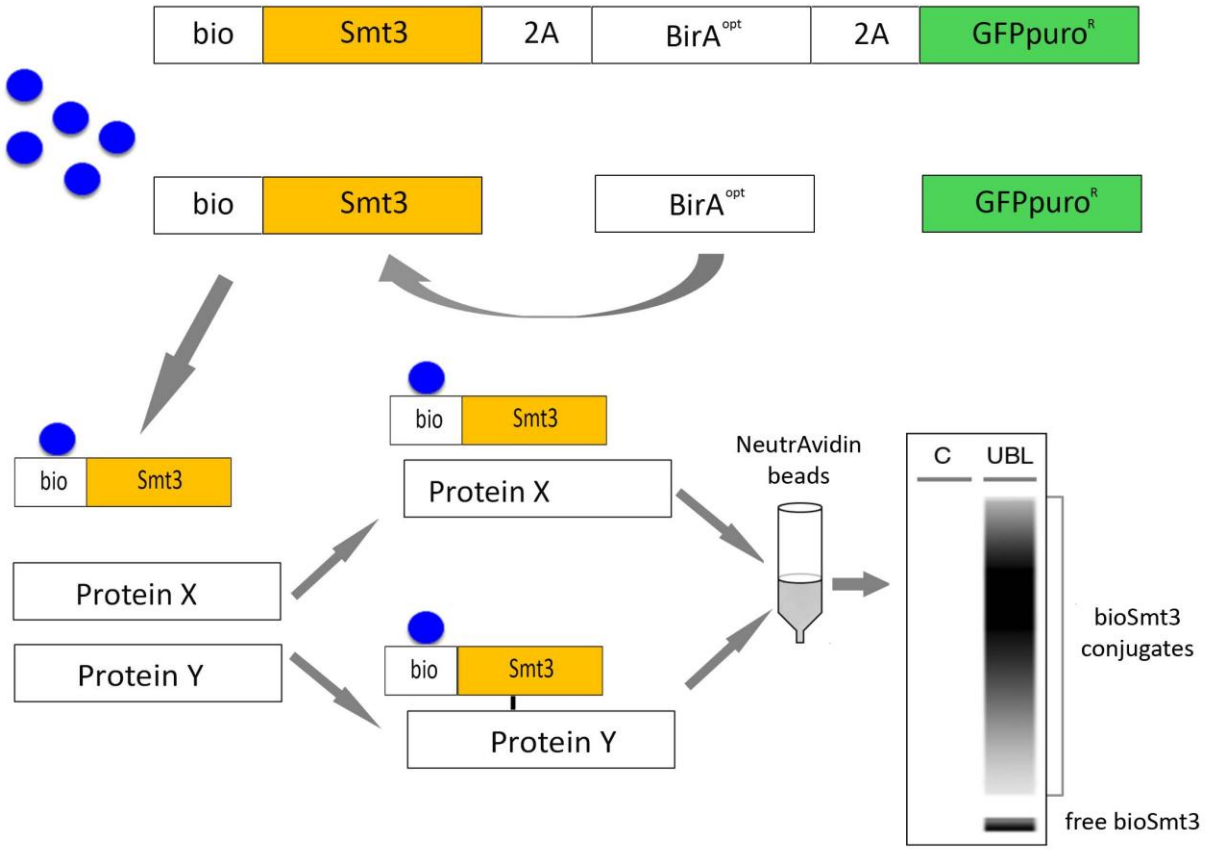


Figure 10. Schematic representation of the bioSmt3 methodology. See text for details.

We transformed S2R+ cells with *pUAST-bioSmt3::BirA*, accompanied or not with *pAc5-Gal4*, and proceeded with the pulldown experiments. However, our results showed that the efficiency of the bioSmt3::BirA processing was approximately 50% (Fig. 11) indicating that, differently than what occurs with the bioUb::BirA fusion (Franco et al., 2011), the endogenous Ulp proteases where not sufficient to process totally the bioSmt3::BirA fusion protein.

We decided to optimize the system by introducing some modifications in the *pUAST-bioSmt3::BirA* vector. First, to avoid the deficient processing of the bioSmt3::BirA fusion

protein by the endogenous Ulps, we used the multicistronic vector *pAc5-STABLE2-Neo* (González et al., 2011). Here, the *bioSmt3* and *BirA* sequences were separated by "T2A" or "2A-like" sequences, also called CHYSEL (cis-acting hydrolase element) peptides, which are viral signals that direct self-cleavage of polyproteins into individual proteins (González et al., 2011). Second, we introduced eukaryote codon optimization for *BirA* to facilitate the translation of this protein in *Drosophila* cells and *in vivo*. This optimized protein was called $BirA^{opt}$.

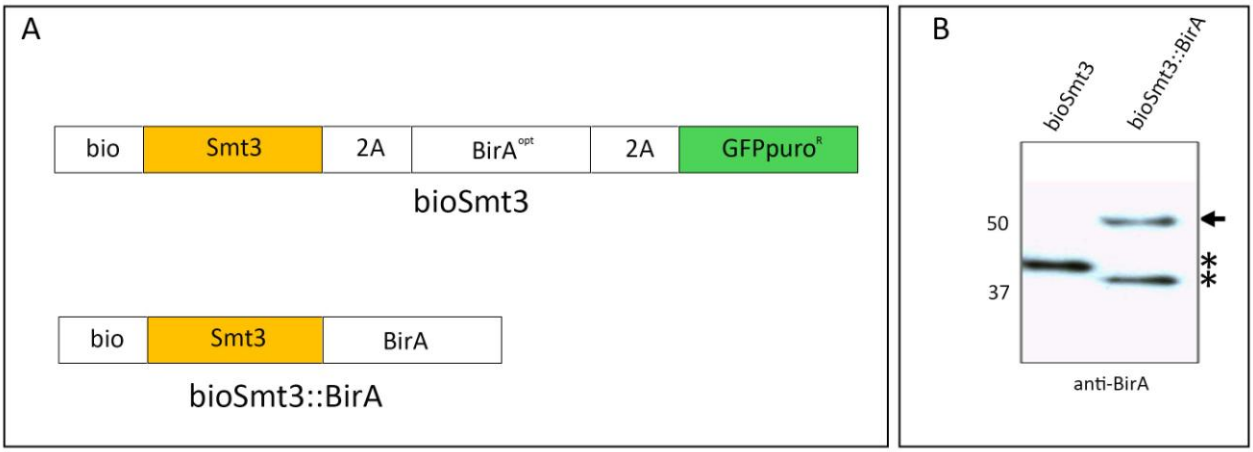


Figure 11. Efficiency comparison between bioSmt3 and bioSmt3::BirA. A. Schematic representation of the plasmids *pAc5-bioSmt3-2A-BirA^{opt}-2A-GFPpuro* (pAc-bioSmt3) and *pUAST-bioSmt3::BirA*. **B.** Anti-BirA Western blot of S2R+ transfected with *pAc5-bioSmt3-2A-BirA^{opt}-2A-GFPpuro* or *pUAST-bioSmt3::BirA*. Black arrow indicates polypeptide bioSmt3::BirA not cleaved (50 KDa) and asterisks indicate free BirA. BirA in the lane of bioSmt3 appears shifted due to the presence of part of the polypeptide 2A. Molecular weight markers are shown to the left.

S2R+ cells were transfected with the new *pAc5-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro* (*pAc5-bioSmt3*) vector or with *pAc5-BirA^{opt}-T2A-GFPpuro* (*pAc5-BirA^{opt}*) vector as a control. GFPpuro was introduced in the last module of the vector as a tool to control the efficiency of cell transfection. The utility to have a selectable marker (puromycin resistant, puro), that allows the generation of stable cell lines, will be treated more extensively in the *Discussion*. After pulldown with the NeutrAvidin beads, we saw a notable enrichment of biotinylated proteins in the sample (Fig. 12). Therefore, the system resulted to be an efficient tool that could be used to isolate SUMOylated protein in *Drosophila* cells.

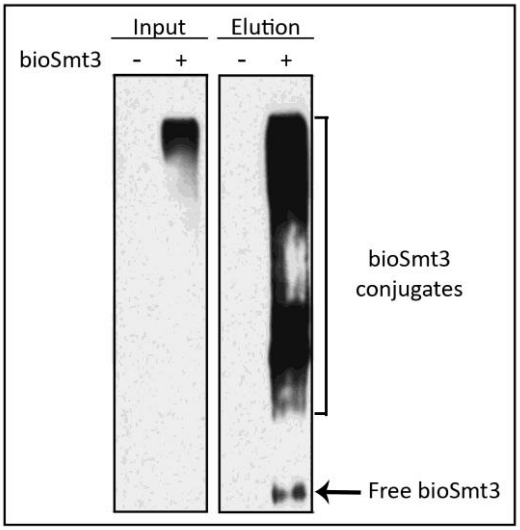


Figure 12. Enrichment of bioSmt3 conjugates using the vector *pAc5-bioSmt3.* Anti-biotin Western blot of S2R+ pulldown. *Drosophila* cells were transfected with *pAc5-bioSmt3* (bioSmt3). The (+) Elution lane shows the enrichment of the bioSmt3 conjugates. Arrow indicates the free bioSmt3.

1.2. Localization of bioSmt3 in *Drosophila* cells

As described in the *Introduction*, Smt3 localization is mainly nuclear and, as in the case of the SUMOs mammalian homologs, Smt3 aggregates in discrete domains (Bhaskar et al., 2000; Lehembre et al., 2000; Smith et al., 2004). To confirm the functionality of the biotin-conjugation system, we analyzed the localization of bioSmt3 in the cells. Our purpose was to check whether the exogenous bioSmt3 could reproduce the subcellular localization of the endogenous protein. We transfected S2R+ cells with *pAc5-bioSmt3* and bioSmt3 were visualized using streptavidin conjugated to a fluorescent dye. To identify the endogenous Smt3, we used anti-Smt3 antibodies in cells transfected with empty *pAc5* vector or in non-transfected cells (Fig. 13). The comparison of the immunofluorescence experiments revealed that bioSmt3 localized in a way similar to the endogenous Smt3. In both cases, the proteins were mainly nuclear and localized in nuclear bodies.

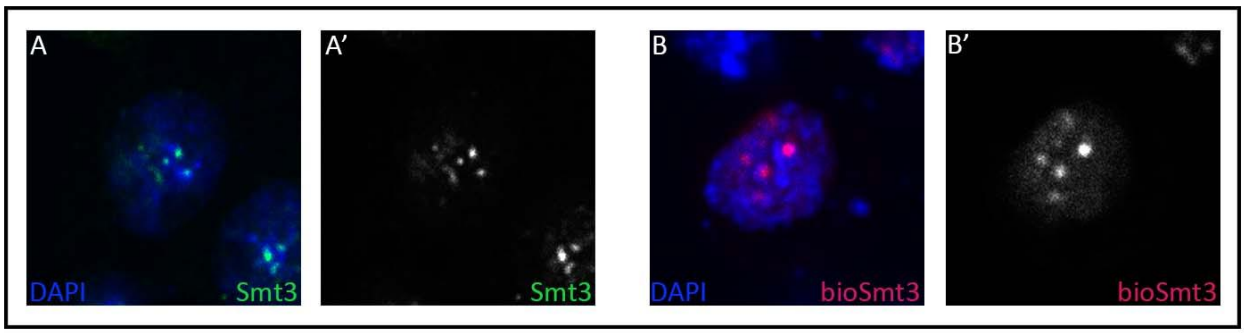


Figure. 13. Localization of Smt3 and bioSmt3. A. Anti-Smt3 antibody was used to visualize endogenous Smt3 (green) in S2R+ cells. **B.** Anti-streptavidin conjugated to Alexa Fluor 594 was used to visualize bioSmt3 (red) in S2R+ cells transfected with *pAc5-bioSmt3.* Nuclei were marked with DAPI (blue). **A', B'.** Green and red channels are shown independently in black and white. Pictures were taken with a confocal microscope.

1.3. SUMOylation of Ftz-f1 by bioSmt3

To analyze the efficiency of our system, we decided to study the SUMOylation of a specific target using the *pAc5-bioSmt3* vector in S2R+ *Drosophila* cells. For this, we chose Fushi tarazu transcription factor 1 (Ftz-f1), a nuclear receptor homologous to the mammalian Steroidogenic factor 1, SF1, which is involved in steroids synthesis in the mammalian adrenal gland (Buaas et al., 2012). As it was published for SF1 (Chen et al., 2004), previous results from the laboratory showed that Ftz-f1 was modified *in vitro* by SUMO1. We used the bioSmt3 system to check the SUMOylation of Ftz-f1 in *Drosophila* cells. S2R+ cells were transfected with *pUASTattB-Flag-βFtz-f1* together with *pAc5-bioSmt3* or *pAc5-BirA*^{opt}. After a NeutrAvidin chromatography column, we visualized the SUMOylated form of Ftz-f1 in a Western blot by using anti-Flag antibodies. As shown in Fig. 14, Ftz-f1 appeared at the expected size in the Input panel in presence or absence of bioSmt3, while in the Elution panel two bands of a higher molecular weight than the non-modified Ftz-f1 appeared only in presence of bioSmt3. We performed the same experiment with the vector *pUAST-bioSmt3::BirA*. When we compared the results with the two vectors, we could confirm that *pAc5-bioSmt3* conferred a higher efficiency of target conjugation than *pUAST-bioSmt3::BirA*. The results on Ftz-f1 SUMOylation were published as part of the work done in the laboratory on the role of SUMO and Ftz-f1 during *Drosophila* steroidogenesis (Talamillo et al., 2013).

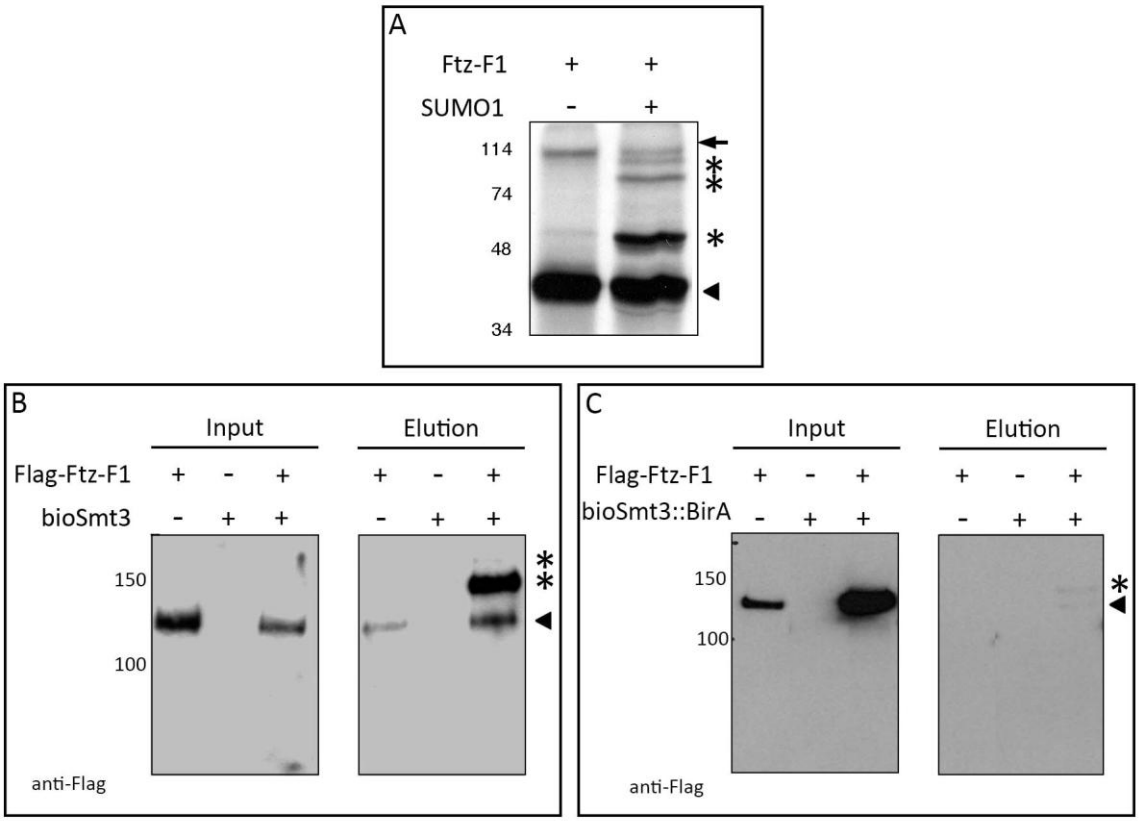


Figure 14. SUMOylation of Ftz-F1 (Talamillo et al., 2013). **A**. *in vitro* SUMOylation of Ftz-F1. In presence of SUMO1 (+) modified form of Ftz-F1 indicated by asterisks. Arrowhead indicate Ftz-F1 non-modified form that appeared in both lane. Arrow indicate nonspecific band. Molecular weight markers are shown to the left. as published in (ref). . **B, C.** Western blot of pulldowns to analyze the SUMOylation of Ftz-F1 in S2R+ cells. In **B** the new vector *pAc5-bioSmt3* was used while, in **C,** *pUAST-bioSmt3::BirA* was used. Asterisks in the Elution panels indicate modified forms of Ftz-F1 that are present in higher proportion in panel **B** than in **C**. Arrowheads indicate Ftz-F1 non-modified form that appeared in the Elution lane, probably due to non-specific interaction to the NeutrAvidin beads. Molecular weight markers are shown to the left.

In view of our results, we concluded that the bioSmt3 system could be applied in cultured cells for the identification of specific SUMOylated proteins. Then, we hypothesized that the system could be used in combination with Mass Spectrometry for the identification of new SUMOylated targets in *Drosophila*.

1.4. Isolation and identification of new bioSmt3 conjugates in *Drosophila* cultured cells

In order to facilitate the Mass Spectrometry large-scale experiments, we introduced an additional improvement to the *pAc5-bioSmt3* vector. As mentioned before, the multicistronic *pAc5-STABLE2-Neo* vector was used as backbone to introduce *bioSmt3* and *BirA^{opt}*. *pAc5-STABLE2* vectors are characterized by having different independent modules that could be easily substituted. To increase the efficiency of the bioSmt3 conjugation we decided to switch the selectable marker *GFPpuro* by the E2 *Drosophila* enzyme *lwr*. At the same time, we introduced ten repetitions of the *UAS* sequence in the *pAc5* promoter, which increased significantly the level of expression of *bioSmt3*. S2R+ cells were transfected with the *pAc510XUAS-bioSmt3-T2A-BirA^{opt}-T2A-lwr* (*pAc510XUAS-bioSmt3-lwr*) vector, or with *pAc5-10xUAS-FlagCherry-T2A-BirA^{opt}-T2A-lwr* (*pAc510XUAS-BirA^{opt}-lwr*) as a control, in presence of *pAc5-Gal4.* After the pulldown performed with NeutrAvidin beads, total protein eluates were resolved on a SDS-PAGE and stained with Colloidal Coomassie (Fig. 15).

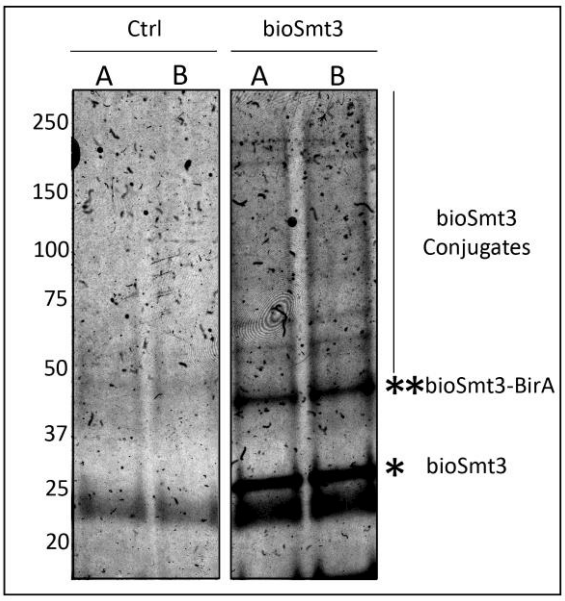


Figure 15. bioSmt3-conjugates in S2R+ *Drosophila* cells. Gel stained with Colloidal Coomassie. A and B correspond to two replicates of pulldown elution performed with S2R+ cells transfected with *pAc510XUAS-BirA^{opt}-lwr* (Ctrl) or *pAc510XUAS-bioSmt3-lwr* (bioSmt3). A single asterisk indicates unconjugated bioSmt3, 2 asterisks indicate the bioSmt3-BirA not processed fusion and the vertical line indicates the high molecular weight bioSmt3 conjugates. Molecular weight markers are shown to the left.

Bands were excised from the gel, subjected to trypsin digestion and analyzed by Mass Spectrometry in collaboration with the group of Dr. Jesper Olsen at the University of

Copenhagen (Denmark). We identified 1341 proteins from this Mass Spectrometry analysis. After our analysis of the intensities ratio between the experimental samples and the controls (explained in *Materials and Methods*), we obtained a list of 980 putative SUMOylated proteins that are enriched in the elution fraction of the bioSmt3 transfected cells (Appendix I).

We performed a bioinformatics analysis of the 980 selected proteins. When comparing our bioSmt3 list with the Mass Spectrometry analysis of SUMOylated proteins performed by Nie and coworkers in *Drosophila* embryos (Nie et al., 2009), we saw that a third of the proteins identified (134) were present in the bioSmt3 list (Fig. 16A). We considered this as a good overlap, taking into account that we were comparing two different biological systems (cells and embryos). The Gene Ontology (GO) analysis revealed that the putative SUMOylation targets in *Drosophila* cells were involved in different biological process that could be grouped in 4 categories: metabolism, development, component organization or localization and transport. The majority of the 980 proteins were localized in the nucleus, distributed as proteins associated to the nucleolus, the nuclear envelope or the nucleoplasm. The portion of proteins present in the cytoplasm was classified in different GO categories, from which the most representative are protein complexes and cytoskeleton (Fig 16B). In respect to the molecular function, most of the proteins identified were factors involved in gene expression regulation. GO categories emerging from the analysis were protein, RNA, DNA or ion binding proteins, as well as proteins involved in DNA catalytic activity (Fig. 16B). Approximately 20% of the bioSmt3-conjugates identified by Mass Spectrometry are involved in transcriptional regulation, which is considered to be the main target group for SUMOylation in the cell. Taken together these results, our analysis coincides with that made by others in *Drosophila* and other systems, reinforcing the validity of our method. To further confirm the targets identified in our Mass Spectrometry analysis, we performed validation experiments using the bioSmt3 system and Western blot analysis.

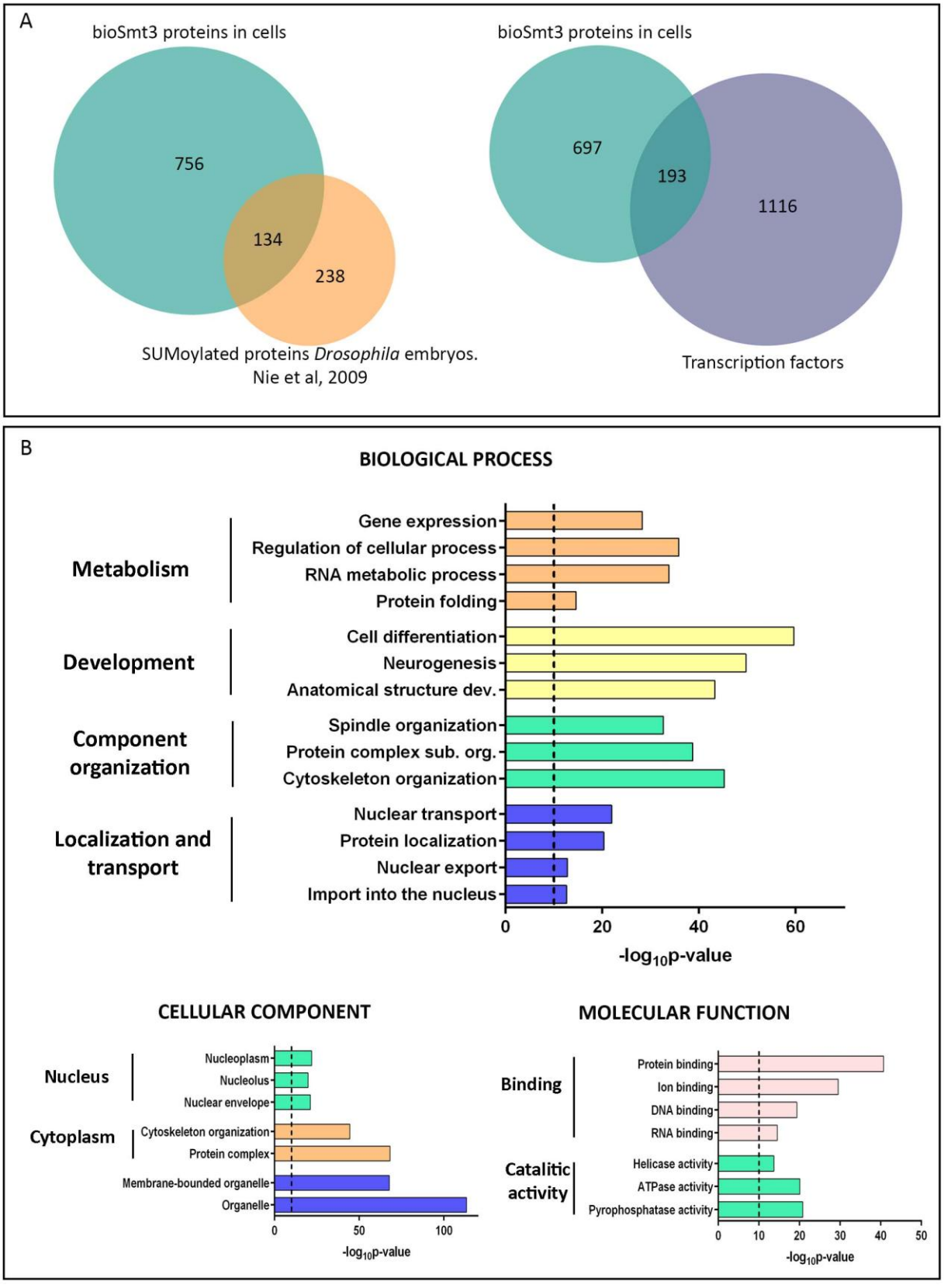


Figure 16. Analysis of bioSmt3-conjugated proteins identified by Mass Spectrometry in S2R+ *Drosophila* cells. A. Comparison between the bioSmt3 list and the list of SUMOylated proteins identified in *Drosophila* embryos (Nie et al., 2009) or with a list of the transcription factors known in *Drosophila.* **B.** GO analysis for biological process, cellular compartment and molecular function performed using Vlad.

1.5. Validation of bioSmt3 conjugates in *Drosophila* cultured cells

For the validation experiments, we firsts chose some targets that we overexpressed in *Drosophila* cells. Three examples of candidates for SUMOylation, that were identified in the Mass Spectrometry analysis and we were able to validate, are shown in Fig.17: the homolog of NF-κB in *Drosophila*, Relish (Rel), and two nuclear envelope proteins, Kugelkern (Kuk) and Otefin (Ote). For these experiments, S2R+ cells were transfected with the mentioned proteins fused to the Flag tag in presence of *pAc510XUAS-bioSmt3-lwr*, or *pAc510XUAS-BirA^{opt}-lwr*, together with *pAc5-Gal4*. After the NeutrAvidin pulldowns, Western blots were developed with anti-Flag antibodies revealing the SUMOylated forms in the Elution panels only in presence of bioSmt3.

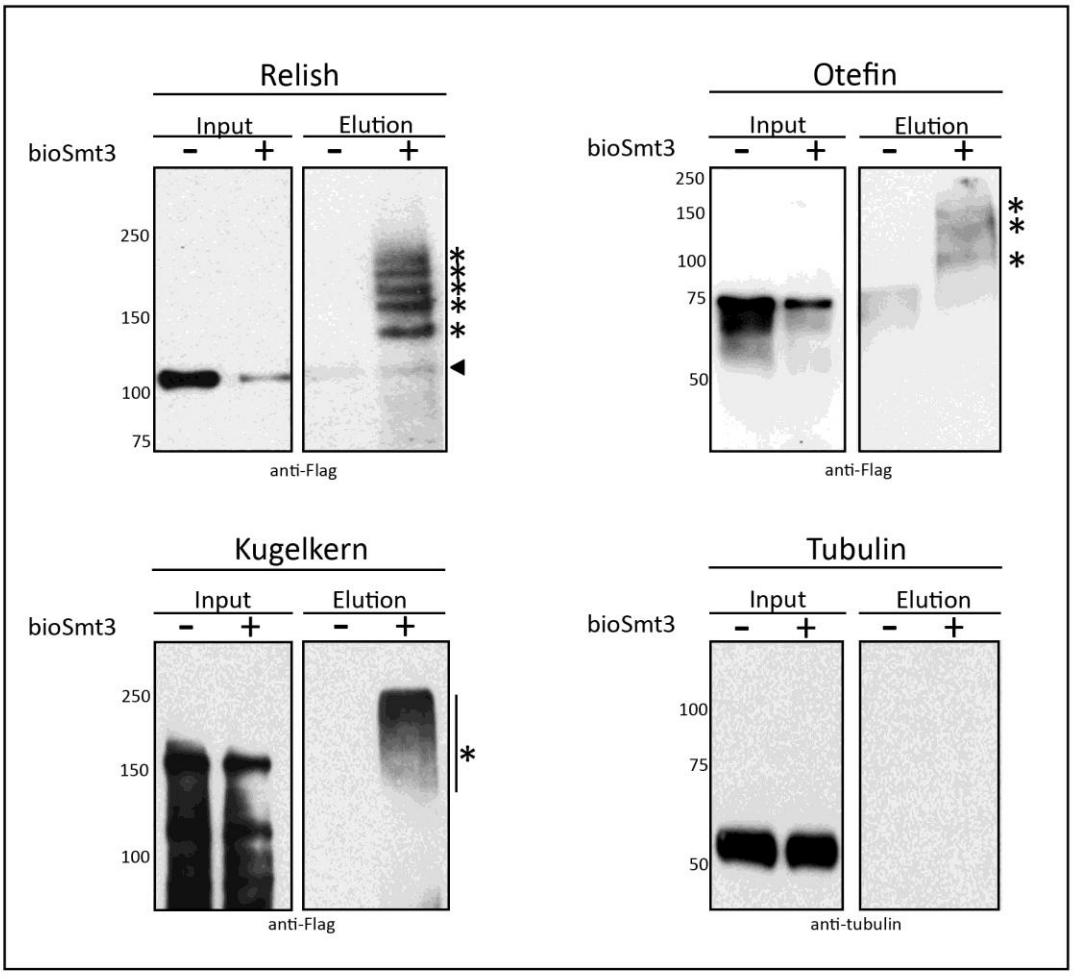


Figure 17. Validation of exogenous SUMOylated proteins. Anti-Flag Western blot of pulldowns performed in S2R+ cells transfected respectively with *Relish*, *Otefin* or *kugelkern* fused to a Flag tag and *pAc510XUAS-bioSmt3-lwr*, (bioSmt3) (+) or *pAc510XUAS-BirA^{opt}-lwr* (-) as a control. In the Elution panels, asterisks indicate the modified forms of each protein. For Relish, the non-modified form of the proteins is indicated with an arrowhead in the Elution panel. Tubulin was used as a loading control. Molecular weight markers are shown to the left.

Despite the small fraction of SUMOylated form in the general pool of each protein, we were able to validate the SUMOylation of some factors by using antibodies against endogenous proteins. This reflects the high specificity and power of the bioSmt3 method. These experiments were done in collaboration with Dr. Wendy Xolalpa in the laboratory. For the validation of the endogenous proteins, we were limited by the antibodies available. We transfected S2R+ cells, as we did for the Mass Spectrometry analysis, with *pAc510XUAS-bioSmt3-lwr* or *pAc510XUAS-BirA^{opt}-lwr* in presence of *pAc5-Gal4*. By using specific antibodies we validated five endogenous targets (Fig. 18). Three of them were previously identified as SUMOylated proteins: Ultraspiracle (Usp), a nuclear receptor involved in steroid signaling (Wang et al., 2014), the transcription factor (Monribot-Villanueva et al., 2013; Nie et al., 2009) and the Eukaryotic initiation factor 4E (eIF-4E) (Xu et al., 2010). The other two validated proteins were not previously described to be SUMOylated in *Drosophila*: the Glutathione S-transferase involved in axonogenesis Failed axon connections (Fax) and the intermediate filament Lamin (Lam). While the mammalian homolog of Lam was demonstrated to be modified by SUMO (Zhang and Sarge, 2008), it is the first time that any Fax homolog is identified as a target of SUMOylation. Further analysis would be required to determine the *in vivo* function of these proteins SUMOylation.

In conclusion, our analysis in cultured cells served as proof of principle for the bioSmt3 system. We decided then to apply this technology *in vivo*.

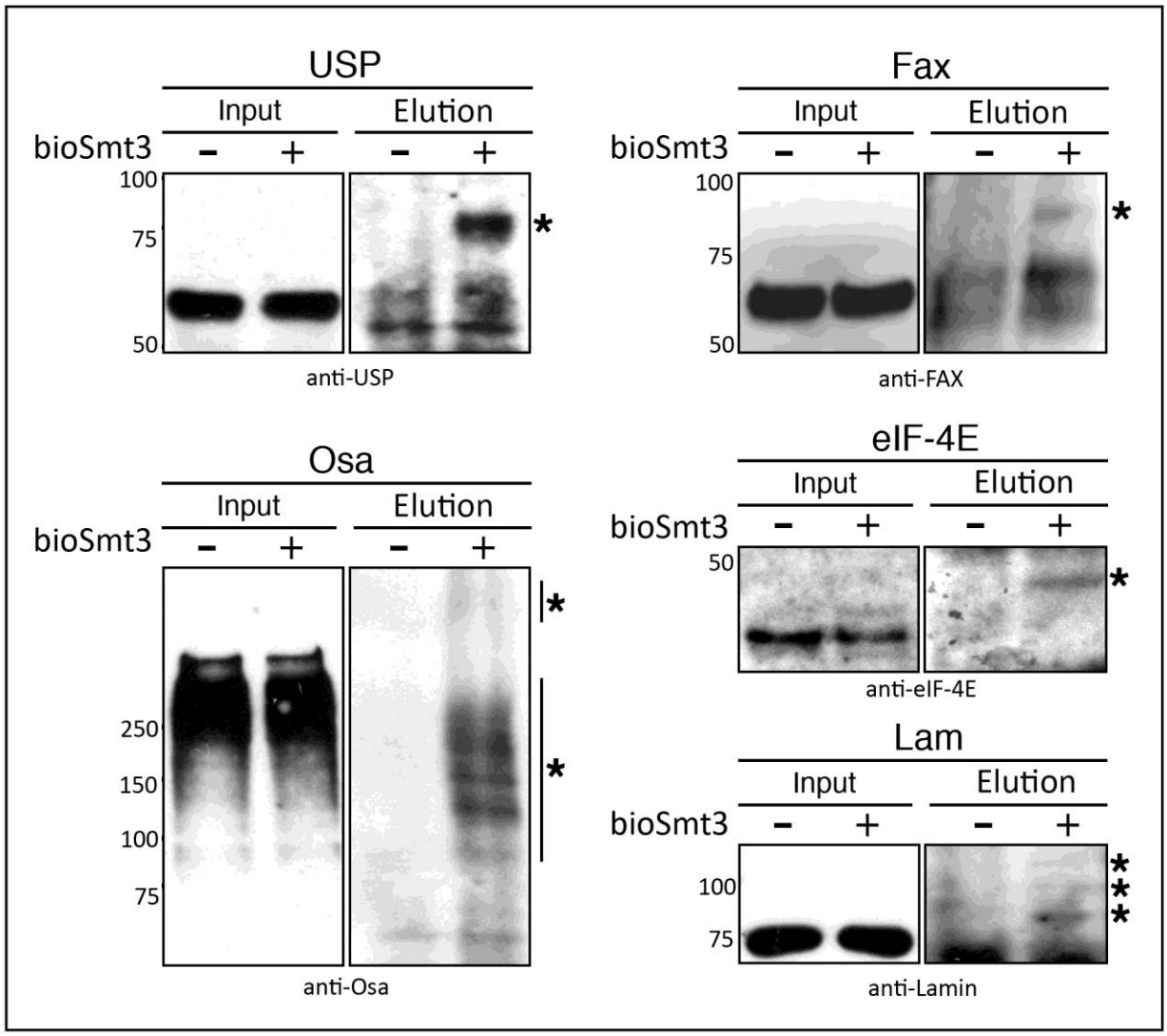


Figure 18. Validation of endogenous SUMOylated proteins. Western blot of pulldowns performed in S2R+ cells. Antibodies against specific endogenous proteins were used. In the Input panels, the non-modified forms of the proteins are shown and in the Elution panels asterisks indicate the modified forms. Molecular weight markers are shown to the left.

1.6. Isolation of bioSmt3 conjugates *in vivo*

As described in the *Introduction*, Smt3 is necessary for steroidogenesis in *Drosophila*. Conditional knockdown of *smt3* in the PG (*phm-Gal4>pUASt-smt3i*) blocks development at the end of L3 due to a reduction in the ecdysone levels (Talamillo et al. 2008). To check the functionality of the *bioSmt3* construct, we analyzed the rescue capacity of the *pUAST-bioSmt3::BirA* transgene in a silenced *smt3i* background. Surprisingly, *phm-Gal4>pUAST-smt3i,pUAST-bioSmt3::BirA* larvae underwent normal metamorphosis and generated fertile adults. This indicates that the *bioSmt3* construct is totally functional, despite the

reduced efficiency of the bioSmt3::BirA fusion processing. Importantly, in those larvae, the majority of Smt3 available in the cells is in the bioSmt3 form (scheme Fig. 19).

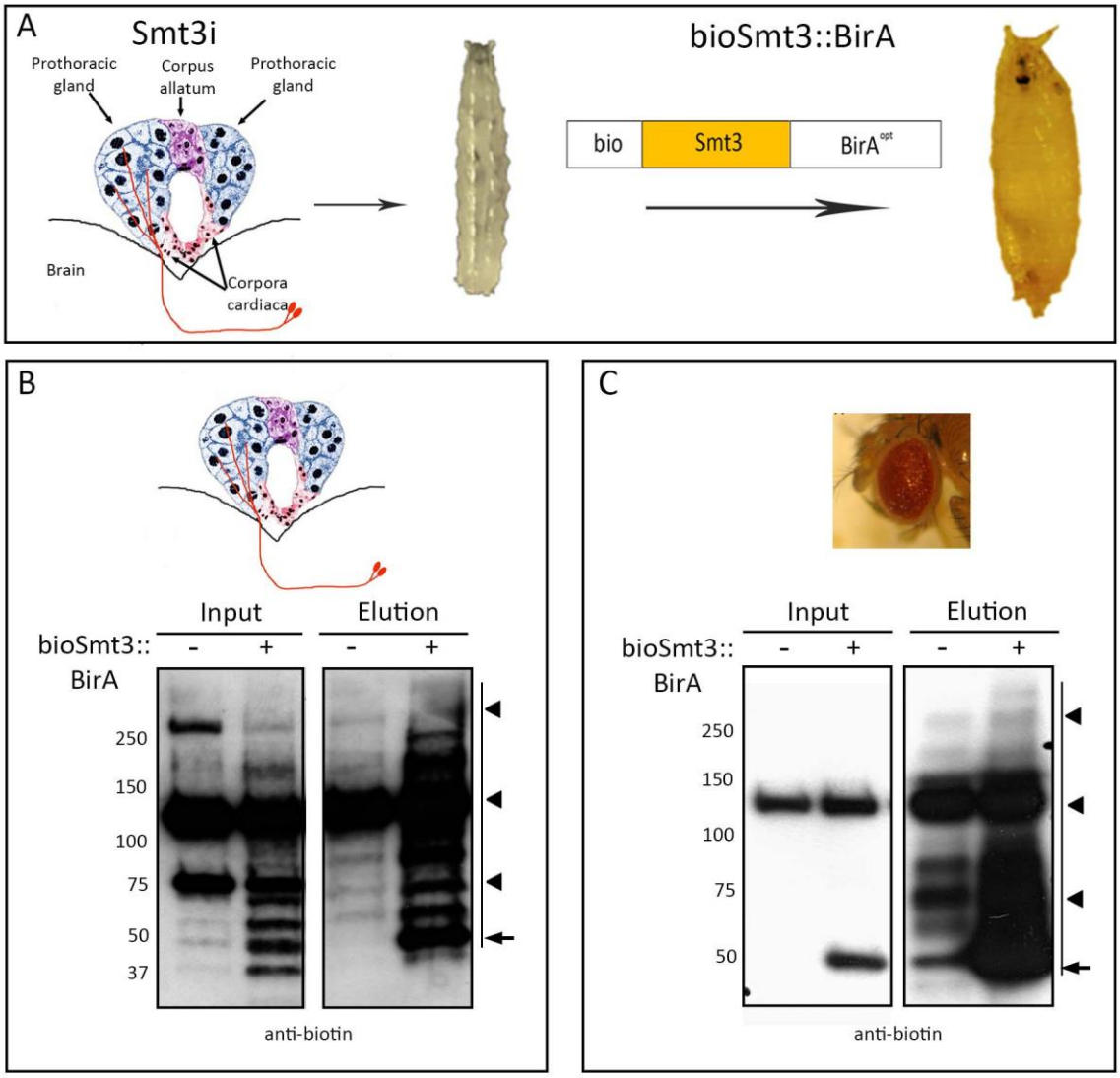


Figure 19. bioSmt3 expression *in vivo*. A. Schematic representation of the functionality of bioSmt3::BirA. Larvae silenced for *smt3* expressing bioSmt3::BirA are able to pupariate**. B, C.** Anti-biotin Western blot of pulldowns performed using animals where bioSmt3 was expressed in the larval PG (**B**) or in the adult heads (**C**) of *Drosophila* silenced for *smt3* and expressing bioSmt3::BirA. The Elution panels (+) show the enrichment of bioSmt3 conjugates. Arrowheads indicate the endogenous biotinylated proteins and arrows the bioSmt3::BirA not processed fusion. Molecular weight markers are shown to the left in **B** and **C**.

In view of these results, we decided to use the *phm-Gal4>UAS-smt3i,UAS-bioSmt3::BirA* larvae for the *in vivo* isolation of SUMOylated proteins, using *phm-Gal4>pUAS-smt3i* as a control. We dissected the brain/ring gland complexes from larvae of these genotypes that were used to perform NeutrAvidin pulldowns. During the Western blot analysis using anti-biotin antibodies, we could appreciate that the amount of biotinylated proteins extracted was not

sufficiently high to recover enough biotinylated proteins for the Mass Spectrometry procedure (Fig. 19). Probably, more tissue was needed to achieve enough material for Mass Spectrometry. However, the small size of the ring gland would make difficult to reach material enough for proteomics and, therefore, we decided to change strategy and isolate the SUMOylated proteins from *GMR-Gal4>UAS-smt3i,UAS,bioSmt3::BirA* flies. *GMR-Gal4* drives expression under the control of *Glass Multiple Reporter* (*GMR*) promoter elements, expressed in the developing eye (Hay et al., 1997), a much larger organ than the ring gland. Adult heads were collected from 10 days-old flies and tissue was used for a NeutrAvidin chromatography and Western blot analysis (Fig. 19). The Western showed an enrichment of the bioSmt3-conjugates in the Elution sample. However, the amount of protein was still not enough for the Mass Spectrometry analysis.

According to Kanakousaki and Gibson (Kanakousaki and Gibson, 2012), SUMOylation occurs mainly in proliferating cells. Proliferation in the eye at adult stages is reduced, which could be the cause of the limited yield of biotinylated proteins in our pulldown experiments. Therefore, we decided to repeat our analysis in the L3 larval stage, where imaginal tissues are proliferating. In addition, we generated new transgenic flies expressing the transgene *pUAST-bioSmt3-T2A-BirA^{opt}*, following the optimization strategy we adopted for cultured cells. To express the transgene in all the imaginal tissues and increase the yield of biotinylated proteins, we used the driver *heat shock* (*hs)-Gal4* which gets activated upon temperature increase. In addition to activate the expression of the driver, the raise in temperature increased the SUMOylation levels in the cells, as SUMO is considered to be a response to various stresses, including heat (Schimmel et al., 2014).

1.7. Localization of bioSmt3 conjugates *in vivo*

Prior Mass Spectrometry, we analyzed the localization of the bioSmt3 transgene *in vivo* in collaboration with Dr. C. Pérez in the laboratory. We detected bioSmt3 using streptavidin labeled with a fluorescent dye in salivary glands from larvae silenced for endogenous *smt3* (*hs-Gal4>UAS-smt3i,UAS-bioSmt3-T2A-BirA^{opt}*) or from control larvae (*hs-Gal4>UAS-GFP*). In addition, we added biotin in the food of the larvae and we performed several rounds of heat shock treatment at 37^{o}C. In a *smt3i* background, with biotin in the food and after heat shock, we observed the accumulation of biotin in nuclear bodies in a pattern similar to the endogenous Smt3 (Fig.20). Therefore, we concluded that those were the suitable conditions for the Mass Spectrometry analysis.

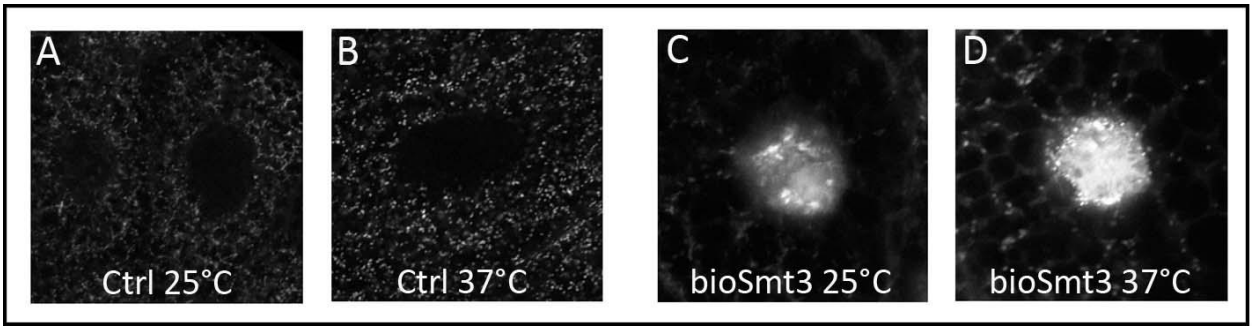


Figure 20. Localization of bioSmt3 *in vivo.* A, B. Salivary glands from control larvae (*hs-Gal4>UAS-GFP*) at 25°C or 37°C. **C, D.** Salivary glands from larvae expressing bioSmt3 (*hs-Gal4>UAS-smt3i,UAS-bioSmt3-T2A-BirA^{opt}*) at 25°C or 37°C. bioSmt3 was maked with streptavidin conjugated to a fluorescent dye.

1.8. Identification of bioSmt3 conjugates *in vivo*

For the Mass Spectrometry analysis of SUMOylated proteins *in vivo,* we performed several rounds of heat shock treatment at 37°C and collected L3 larvae of *hs-Gal4>UAS-Smt3i,UAS-bioSmt3-T2A-BirA^{opt}* or *hs-Gal4>UAS-Smt3i,UAS-BirA^{opt}* genotypes. To reduce the proportion of non-proliferating tissues, we dissected the anterior part of the larval body where most of the imaginal discs are located. After pulldown performed with NeutrAvidin beads, we checked the enrichment of bioSmt3 conjugates (Fig 21A). Protein eluates were resolved on a SDS-PAGE and stained with Colloidal Coomassie (Fig. 21B). Bands were excised from the gel, subjected to trypsin digestion and analyzed by Mass Spectrometry in collaboration with the group of Dr. Jesper Olsen at the University of Copenhagen. The selection of positive targets was done in the same way than for the cell culture analysis (*Results 1.4, Part I*). We identified 92 proteins as putative SUMOylation targets (Appendix II), which were selected based on a 4 fold increased intensity when compared with the control.

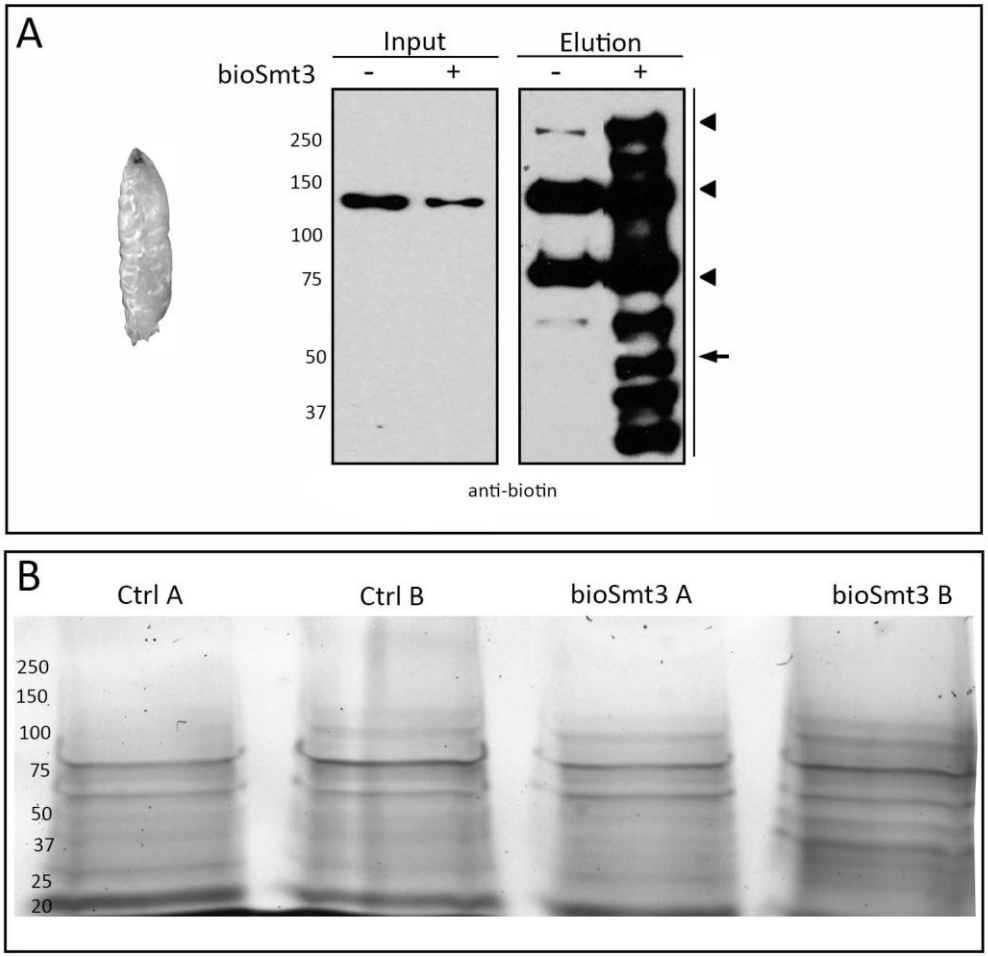


Figure 21. bioSmt3 conjugates *in vivo*. A. Anti-biotin Western blot of larval pulldowns. The Elution panel lane (+) shows the enrichment of bioSmt3-conjugates compared with the control (-). Arrowheads indicate endogenous biotinylated proteins and the arrow indicates bioSmt3::BirA not processed form. **B.** Gel stained with Colloidal Coomassie. A and B correspond to two replicates of the pulldown elution performed with *hs-Gal4>UAS-Smt3i,UAS-BirA^{opt}* (Ctrl) or *hs-Gal4>UAS-Smt3i,UAS-bioSmt3-T2A-BirA^{opt}* (bioSmt3). Molecular weight markers are shown to the left.

Our bioinformatics analysis revealed that approximately 60% of the proteins identified *in vivo* were present in the list of SUMOylated proteins obtained from cultured cells and 40% were present in the list of SUMOylated protein identified by Courey and coworkers in *Drosophila* embryos (Fig. 22A) (Nie et al., 2009). We could consider that those sets of data have a good overlap, considering the different developmental stages in which the *in vivo* experiments were performed. The GO analysis revealed that the putative SUMOylation targets in *Drosophila* were involved in similar biological process that we identified for the list of proteins identified in S2R+ cells: component organization, cellular process or localization and transport. In this case, there was a higher proportion of proteins localized in the cytoplasm, the category of lipid particle, protein complexes and microtubule cytoskeleton being strongly represented. In respect to the molecular function, as we observed in S2R+ cells, the proteins identified were involved in

nucleotide or GTP binding, or in pyrophosphatase or GTPase catalytic activity, which are linked to gene expression regulation and signaling (Fig. 22B). Another GO category emerging from the analysis was structural constituent of cytoskeleton (Fig. 22B). Our analysis *in vivo* let to the identification of less targets than the analysis in cells and further studies should optimize the protein extraction and pulldown from the tissues *in vivo*. Nevertheless, also in this case, our analysis coincides with that

made by others in *Drosophila* or other systems, reinforcing the validity of our method.

Taken together, our results showed that the bioSmt3 system could be efficiently applied in cultured cells and *in vivo* in *Drosophila* to identify SUMOylated proteins. Based on this, and taking into consideration the modularity of the vectors, we wondered whether a similar approach could be applied to mammalian cells for the identification of SUMOylated proteins.

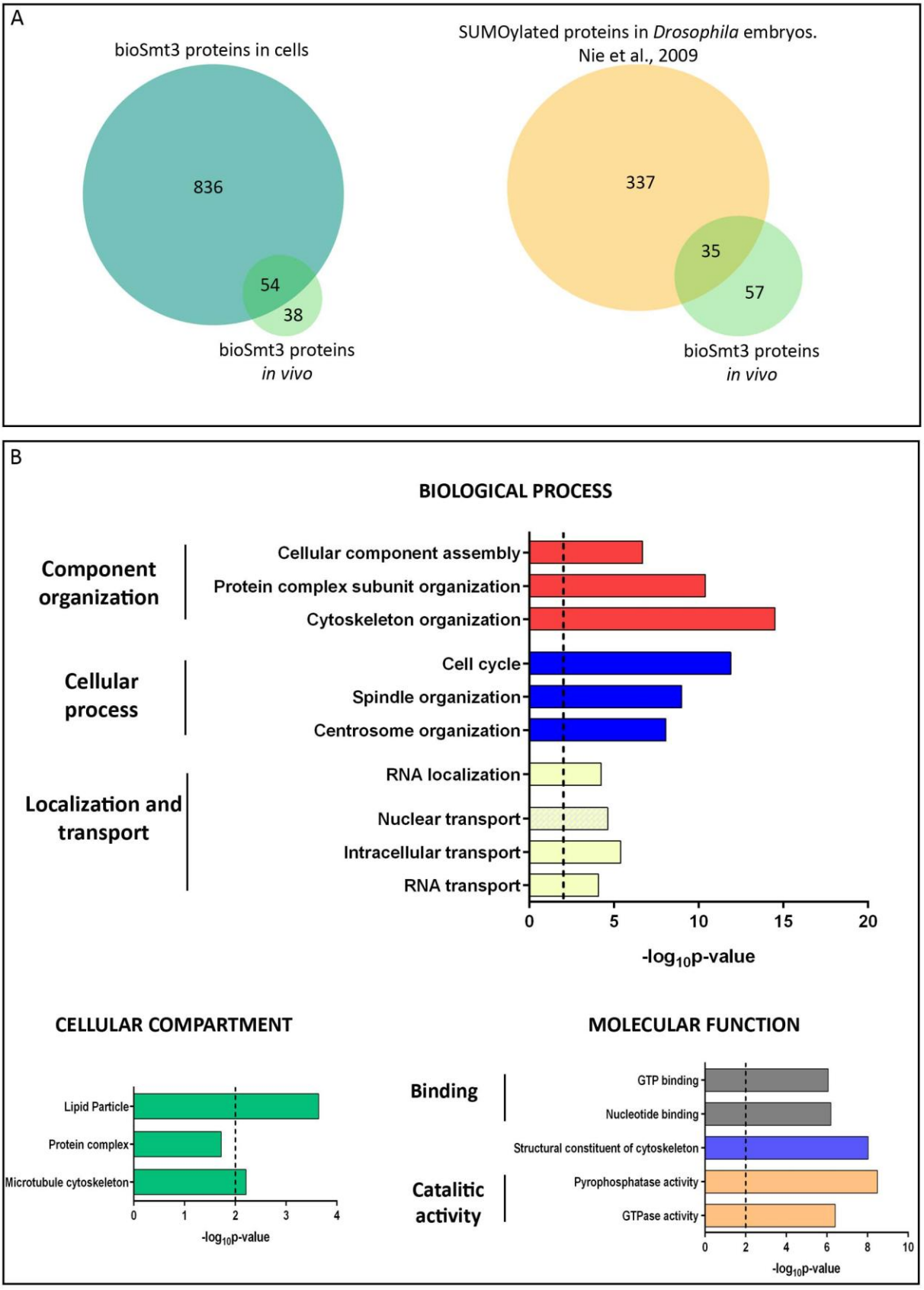


Figure 22. Analysis of bioSmt3-conjugated proteins identified by Mass Spectrometry *in vivo* A. Comparison between the list of bioSmt3-conjugated proteins *in vivo* and the bioSmt3-conjugated proteins in cells (left) or the list of SUMOylated proteins identified in *Drosophila* embryos (right; Nie et al., 2009). **B.** GO analysis for biological process, cellular compartment and molecular function was performed using Vlad.

2. Development of bioUbiquitin-like (bioUbL) vectors for the analysis of post-translational modifications in mammalian cells

2.1. New methodology for the isolation of SUMOylated proteins in mammalian cells

The efficiency and the characteristics of the bioSmt3 system allowed us to apply the same approach to mammalian cultured cells. This part of the project was also done in collaboration with Dr. J. D. Sutherland in the laboratory. This new methodology could be used in different cultured cell systems to study diverse post-translational modifications. First, the modularity of the *pAc5-STABLE2-Neo* vector facilitated the switch of the promoter specificity from insects (*Ac5*) to mammalian (*CAG*) (González et al., 2011). Second, we substituted the *Drosophila smt3* with mammalian SUMO1 or SUMO3 moieties. As we did for the *Drosophila* cells, we chose a specific target to test the efficacy of the system in mammalian cells. We chose the transcription factor SALL1 as target, which, as mentioned in the *Introduction*, it was previously described to be SUMOylated *in vitro*. The experiments performed to study the SUMOylation on SALL1 will be detailed in the *Part II* of the *Results*. Briefly, we were able to confirm the SUMOylation of SALL1, as well as its partial colocalization with SUMO, by using the bioSUMO constructs (Figs. 29,32). These results confirmed the suitability of the system also in mammalian cells.

2.2. Isolation and identification of SUMO3 subproteome

Once we demonstrated that the bioSUMO/BirA system was effective for the isolation of a particular SUMOylated protein in mammalian cells, we proceed to analyze whether the system was suitable for Mass Spectrometry analysis. For that, we transfected HEK 293FT cells with the vector *CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro* or the control *CAG-BirA^{opt}-T2A-GFPpuro,* passed them through a NeutrAvidin column and analyzed them by SDS-PAGE. The Western blot analysis using anti-biotin antibodies showed the isolation of SUMO3-SUMOylated proteins in an appropriated yield to perform Mass Spectrometry (Fig. 23). Mass Spectrometry analysis was done by the Proteomics Platform at CIC bioGUNE in collaboration with Dr. F. Elortza. In the list of candidates to be SUMOylated by bioSUMO3 (Appendix III), we found

proteins that were previously described as SUMOylated, demonstrating the validity of the system.

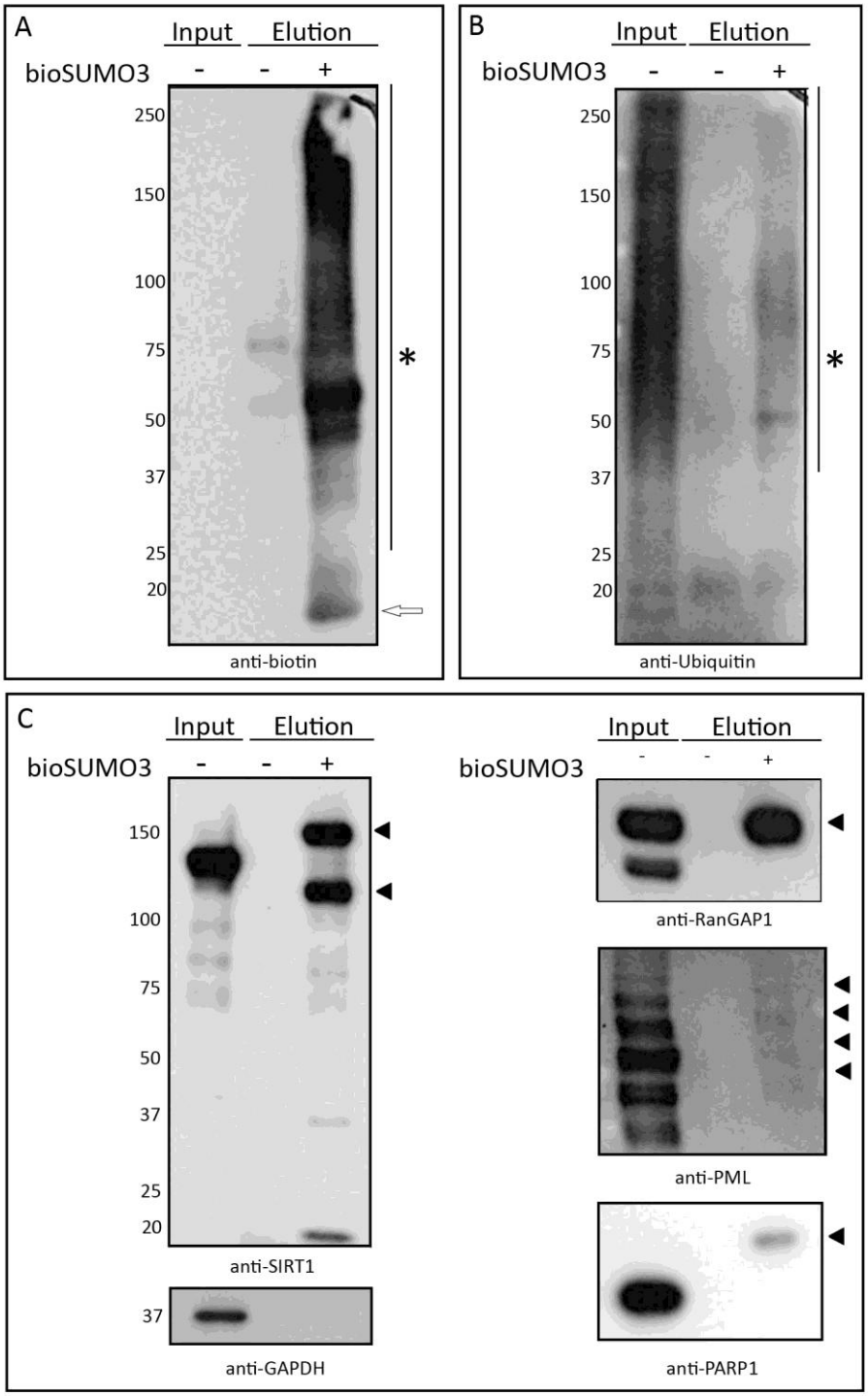


Figure. 23. Identification of bioSUMO3 conjugates and validation. A. Anti-biotin Western blot of pulldowns performed with HEK 293FT cells transfected with *CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro* (+) or *CAG-BirA^{opt}-T2A-GFPpuro* as control (-). In the Elution lane with bioSUMO3 (+), conjugates were enriched (asterisk). The arrow indicates uncojugated bioSUMO3. **B.** Anti-Ub hybridization of the same Western blot shown in **A** after stripping, revealing the presence of Ub-conjugates among the SUMO3-modified isolates (asterisk). **C.** Validation of SUMO3 targets. Different antibodies were used against endogenous proteins. Arrowheads indicate the modified proteins in the Elution lanes. Anti-GAPDH was used as a loading control. Molecular weight markers are shown to the left.

To validate the results, we performed pulldowns in a similar way than the ones done for the Mass Spectrometry analysis. These experiments were done in collaboration with A. Ruiz de Sabando and Dr. J. D. Sutherland in the laboratory. By Western blot using specific antibodies, we validated the SUMOylation by SUMO3 of several endogenous proteins such as Sirtuin 1 (SIRT1), Ran GTPase-activating protein 1 (RANGAP1), Promyelocytic leukemia protein (PML) and Poly [ADP-ribose] polymerase 1 (PARP), all of them previously identified as SUMO targets (Fig. 23). GAPDH was used as loading control in the Input panels and a negative control for the Elution panels, as this protein did not appear in the Mass Spectrometry list.

These experiments demonstrated that the bioSUMO3 system was sufficiently sensitive to detect the SUMOylated fraction of these proteins without overexpressing them and without heat shocking the cells. Interestingly, we identified Ub as one of the targets of bioSUMO3. Furthermore, we detected Ub in the pulldown lane of the Western blot, which is compatible with the existence of SUMO3-Ub mixed chains (Fig. 23).

Taken together, these results confirmed that the bioUbL method can be used in mammalian cells to check the modification of a particular protein, either endogenous or overexpressed, as well as to perform high throughput proteomics.

2.3. New methodology for the isolation of proteins modified by other UbLs in mammalian cells

The positive results mentioned in the previous section encouraged us to expand our repertoire of vectors by substituting the bioSUMO3 sequences by other UbLs, including those that are more studied like Ub or Nedd8 and those that are less studied like ISG15, FAT10 or UFM1. This, as I described before, could be easily done due to the modular nature of the multicistronic vectors (Fig. 24A).

We first analyzed the expression and incorporation of the different bioUbLs to target proteins in HEK 293FT cells. The same number of cells were transfected in parallel with the different bioUbL constructs, cells extracts were passed through an NeutrAvidin column and analyzed by Western blot (Fig. 24B). The blot showed that all the UbLs were incorporated to target proteins, although the relative amount of proteins conjugated to the different UbL modifiers varied, as revealed with anti-biotin antibodies. For instance, bioUb and bioNEDD8 showed a greater enrichment than the bioSUMOs, and those a greater enrichment than FAU or UFM1.

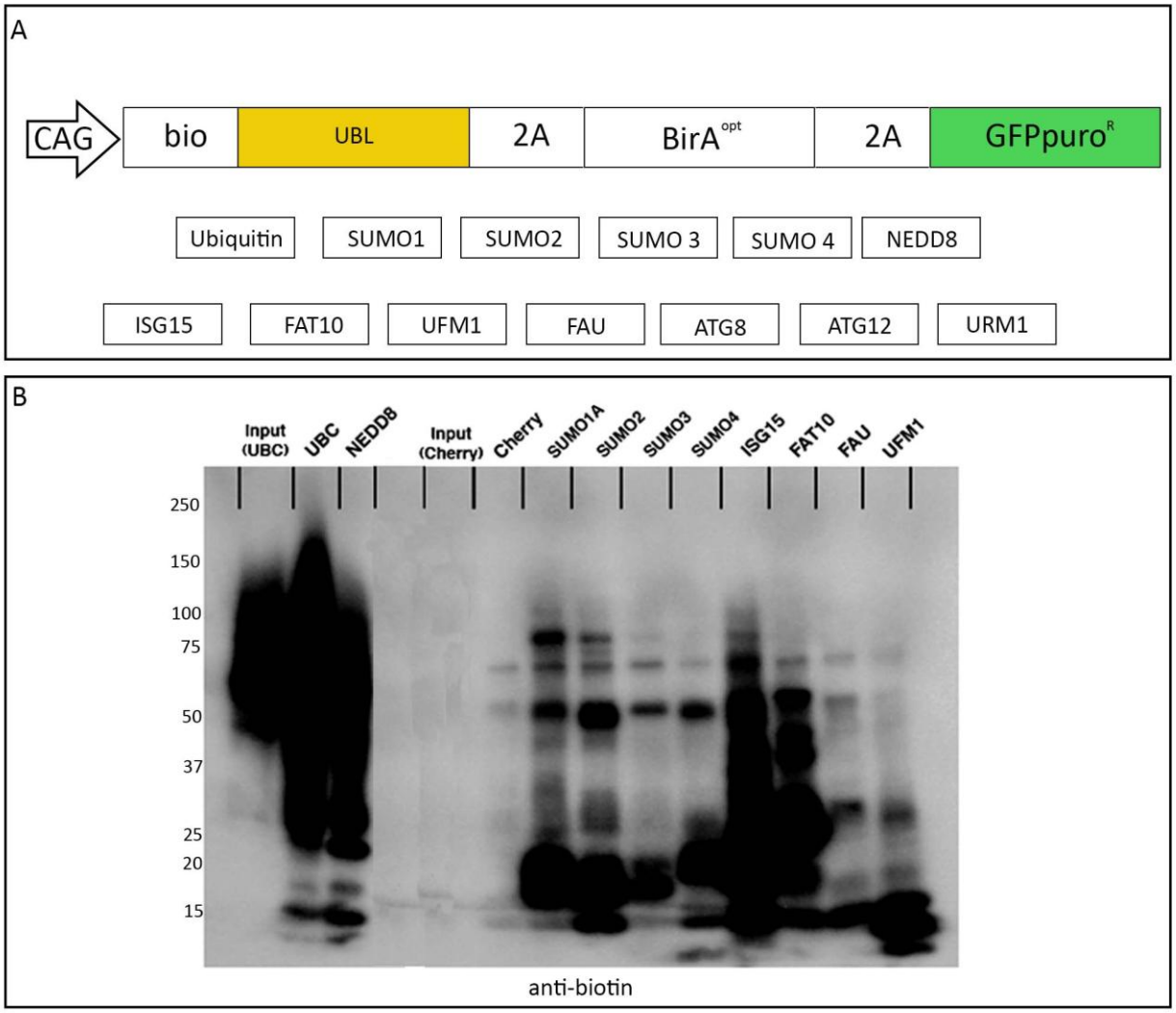


Figure 24. Isolation of bioUbLs-conjugates in mammalian cells. A. Schematic representation of the bioUbL plasmid collection. **B.** Anti-biotin Western blot of pulldowns performed in parallel using HEK 293FT mammalian cells transfected with plasmids expressing the different UbLs. Molecular weight markers are shown to the left.

We then analyzed the cellular distribution of the different bioUbLs by using NeutrAvidin conjugated to a fluorofore. The different bioUbLs showed a distinct pattern of expression, which suggested a distribution driven by the localization of the endogenous target proteins. Most of the UbLs are mainly nuclear although not exclusively, as for example the cases of bioUb, bioNEDD8 or bioFAT10 (Fig. 25).

The heterogeneous distribution of the bioUbLs and their characteristic conjugation pattern suggested that each UbL was incorporated to a specific set of target proteins, supporting the specificity of the bioUbL system.

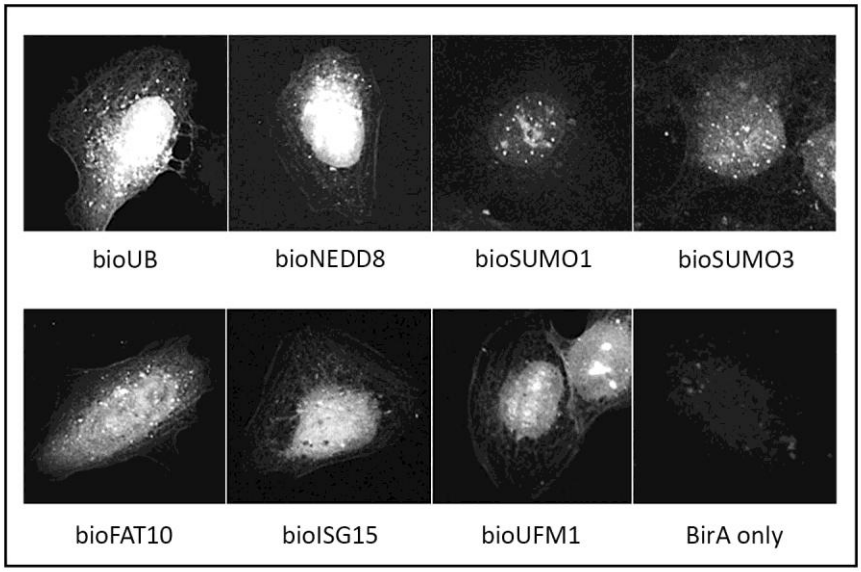


Figure 25. Cellular distribution of different bioUbLs. U2OS cells transfected with different plasmids expressing *bioUbLs* or *BirA* only as a control. Stainings were performed with streptavidin conjugated to Alexa Fluor 594. Pictures were taken with Upright Fluorescent Microscope Axioimager D1.

2.4. Validation of the bioUFM1 system using a known target

To further demonstrate the suitability of the bioUbL system, we chose one of the less studied UbL members, UFM1. To increase the efficiency of UFM1-conjugation, we substituted the last module of *CAG-bioUFM1-T2A-BirA^{opt}-T2A-GFPpuro* by UFC1, the specific E2 enzyme involved in UFMylation.

To validate our approach of bioUbLs as tool to isolate and identified modified targets, we decided to validate a known UFMylated target, as we did previously with the SUMO targets. We chose DDRGK1 (UFBP1) shown to be modified by UFM by two different laboratories. (Komatsu et al. 2004, Lemaire et al. 2011). DDRGK1 interacts with the E3 UFM1-protein ligase, UFL1, and one of its substrates, TRIP4, and is required for TRIP4 UFMylation. Through TRIP4, UFMylation might regulate the transcription mediated by nuclear receptors (Yoo et al., 2014). DDRGK1 could play a role in NF-kappa-B-mediated transcription through regulation of the phosphorylation and the degradation of NFKBIB, the inhibitor of NF-kappa-B (Xi et al., 2013). It could also be involved in the cellular response to endoplasmic reticulum stress, as reported previously for the UFMylation process. Fig. 26 shows the modification of DDRGK when is overexpressed in presence of bioUFM1, but not in presence of BirA^{opt} alone. The modified forms of the protein, which are shifted compared with the non modified protein, could be observed in the Elution panel. These results showed the validity of the bioUFM1

method. We decided then to apply it to the discovery of new UFM1 targets by Mass Spectrometry.

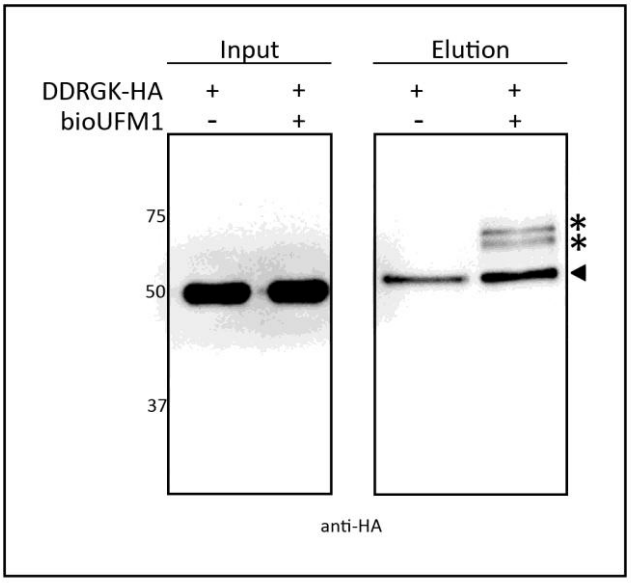


Figure 26. Validation of a UFMylation target. Anti-HA Western blot of a HEK 293FT cells pulldown. Cells were transfected with DDRGK-HA together with CAG-*bioUFM1-UFC1* (bioUFM1) or CAG-*BirA*^{opt} as control. In the Elution lanes, asterisks indicate the modified forms of the protein and the arrowhead indicates the non-modified protein. Molecular weight markers are shown to the left.

2.5. Isolation and identification of UFM1 conjugates

To further demonstrate the suitability of the bioUFM1 system for the high throughput proteomic analysis, HEK 293FT cells were transfected with *CAG-bioUFM1-T2A-BirA*^{opt}*-T2A-UFC1 (CAG-bioUFM1-UFC1)* or *CAG-BirA*^{opt}*-T2A-puro (CAG-BirA*^{opt}*)* as control. After pulldown performed with NeutrAvidin beads, total protein eluates were resolved on a SDS-PAGE and stained with Colloidal Coomassie (Fig. 27).

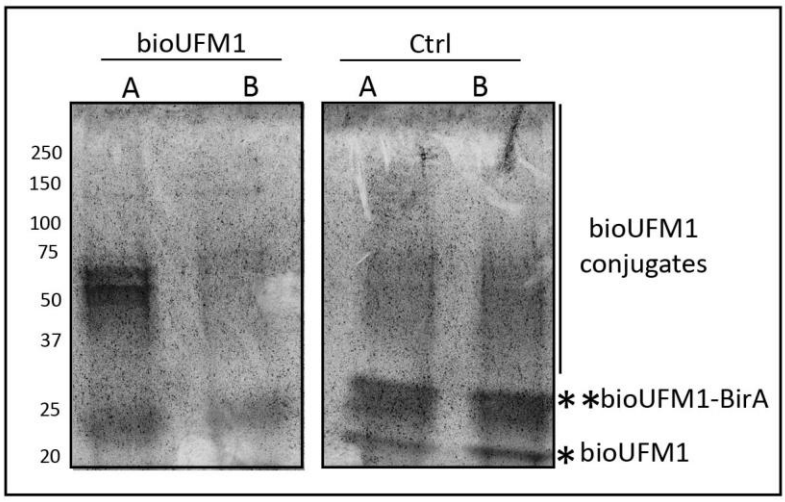


Figure 27. bioUFM1-conjugates in HEK 293FT cells. Gel stained with Colloidal Coomassie. A and B correspond to two replicates of pulldown elutions performed with cells transfected with CAG-*BirA*^{opt} (Ctrl) or *CAG-bioUFM1-UFC1* (bioUFM1). A single asterisk indicates the unconjugated bioUFM1, 2 asterisks indicate the not processed bioUFM1-BirA and the vertical line indicates the bioUFM1 conjugates. Molecular weight markers are shown to the left.

Bands were excised from the gel, subjected to trypsin digestion and analyzed by Mass Spectrometry, in collaboration with the group of Dr. J. Olsen (UCPH, Copenhagen, Denmark). We identified 732 proteins. After our analysis (explained in *Materials and Methods*), we identified 60 proteins as targets of UFM-conjugation (Appendix IV). GO analysis showed that there was a significant enrichment in the biological process categories of translation and mitotic cell cycle. Proteins identified localized in the nucleus and in the cytoplasm (as shown in Figure. 28). In respect to the molecular function the most representative categories from GO analysis are nucleotide binding, in particular to mRNA5'-UTR. These results placed UFM1 in the gene expression regulation function, especially in mRNA and translation regulation.

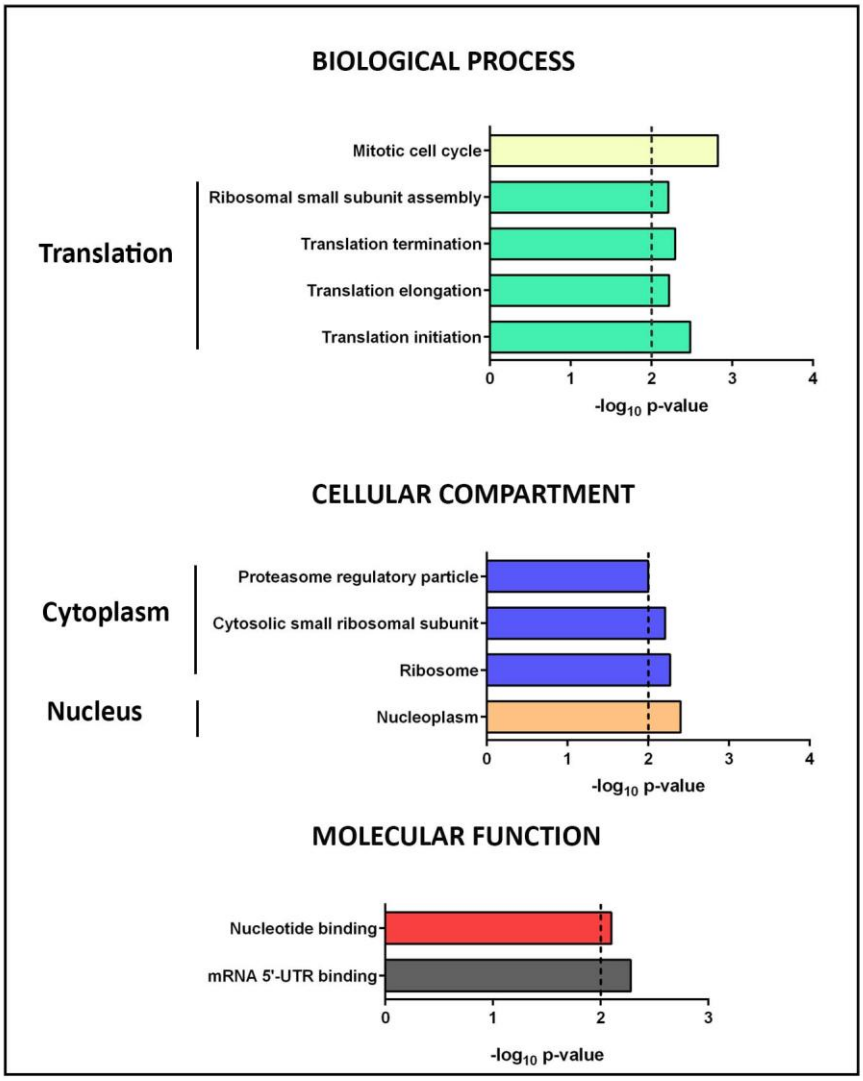


Figure 28. Analysis of bioUFM1 conjugates by Mass Spectrometry. A. GO analysis for biological process, cellular compartment and molecular function performed using innate DB.

Taken together the results shown in the Part I of this thesis, we can conclude that we developed a new efficient system to facilitate the study of SUMOylated proteins in *Drosophila*, in cultured cells and *in vivo*, either for Mass Spectrometry or for the detail analysis of a specific target. Also, we generated a versatile toolbox to facilitate the study of Ub and UbL subproteomes in mammalian cells.

PART II: ANALYSIS OF THE SUMOYLATION OF SALL PROTEINS

1. SUMOylation and localization of human SALL1

1.1. Human SALL1 is SUMOylated in cultured cells

As mentioned in the *Introduction*, human SALL1 is modified post-translationally by SUMO1 *in vitro* (Netzer et al., 2002). Since our long term aim is to explore the role of SUMOylation on SALL protein's function *in vivo*, we decided to check if this transcription factor can be modified by SUMO also in cultured cells. HEK 293FT cells were chosen to check the SUMOylation of SALL1 by using the bioSUMO system described in the *Results, Part I* section.

SALL1 fused to YFP in the C-terminal part was overexpressed in presence of bioSUMO1, bioSUMO3 or the control vector BirA. These SUMO variants, that can be expressed and biotinylated in cells, facilitated the isolation of SUMOylated SALL1 under stringent condition using NeutrAvidin beads. We showed that SALL1-YFP can be modified by both bioSUMO1 and bioSUMO3, corroborating the results published *in vitro* using SUMO1 (Netzer et al., 2002). In Fig. 29A, the Western blot shows the identification of the SUMOylated forms of human SALL1-YFP using anti-GFP antibodies. Higher molecular weight bands were identified corresponding to the modified forms of the protein (Elution panel, asterisk), bigger in size than the non-modified SALL1-YFP (Input panel). As described in the *Introduction*, SALL1 can be SUMOylated *in vitro* at the lysine 1086, localized in the SUMO consensus site between zinc finger domains ZF4 and ZF5 (Fig. 29C): substitution of that lysine by an arginine abolished SUMOylation of a truncated form of SALL1, as it was shown by Netzer and coworkers *in vitro* (Netzer et al. 2002). However, those experiments were done using a truncated form of SALL1 and we cannot discard the presence of other SUMOylation sites.

In order to generate a SALL1 SUMO mutant to be used in our experiments in cultured cells using the bioSUMO system, we analyzed the full length amino acid sequence of SALL1 by the SUMOplot program (http://www.abgent.com/sumoplot). The program suggested 7 possible SUMO consensus sites with a high probability score, that is, more than 0.5 in a range

between 0 to 1. To test the hypothetical SUMOylation consensus sites, we decided to take into account sequences with a score higher than 0.9. Only four sequences respected this very stringent condition and we mutated all of them by substituting each lysine (K571, K592, K982 and K1086) with an arginine (SALL1-4KR; Fig. 29C). As predicted, the SALL1-4KR mutant lost the capacity to be SUMOylated in cells, as shown in Fig. 29B. Therefore, we considered SALL1-4KR a SUMO-null mutant of SALL1.

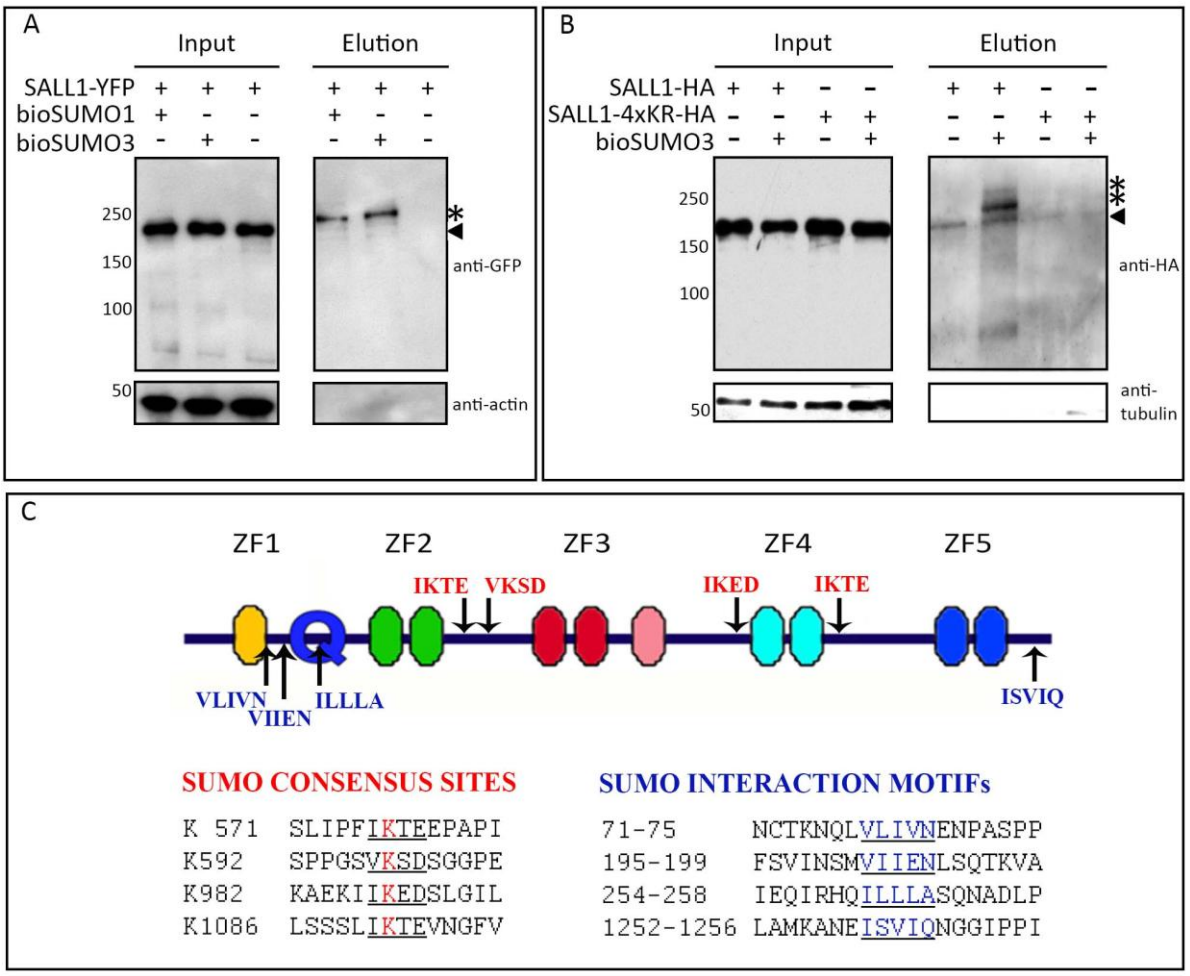


Figure 29. SUMOylation of SALL1 in cultures cells. Western blot of pulldowns to check SUMOylation of SALL1 in HEK 293FT cells. **A.** SALL1 fused to YFP tag was SUMOylated in presence (+) of bioSUMO1 or bioSUMO3. Asterisk in the Elution panel indicates the modified SALL1 that is shifted if compared with the size of non modified SALL1 in the Input panel. Anti-actin antibodies were used as a loading control. Molecular weight markers are shown to the left. **B.** SALL1-4KR fused to HA tag is not SUMOylated in presence of bioSUMO3. Arrowhead indicates a non modified form of SALL1 and the asterisks the SUMOylated SALL1-HA that appears only in the second lane of the blot corresponding to the WT form. In the Input the expression of WT and SUMO mutant of SALL1 are shown. Anti-Tubulin antibodies were used as a loading control. Molecular weight markers are shown to the left. **C.** SALL1 schematic representation. Ovals represent the zinc fingers (ZF) distributed along the protein. Q represents the poly-glutamine domain. In red, SUMO consensus sites mutated in SALL1-4KR and in blue the predicted SIMs.

1.2. SALL1 localizes in nuclear bodies in human cells

Localization of SALL1 and its homologs in nuclear bodies has been reported in cultured cells and *in vivo* (Abedin et al., 2011; Kiefer et al., 2002; Netzer et al., 2001; Sánchez et al., 2010). Our aim was to visualize in cultured cells the correlation between SALL1 and the SUMO proteins. First, we tested our system by studying the subcellular localization of human SALL1 in different cell types. For the immunofluorescence experiments we decided to use human U2OS cell line, which expresses SALL1 endogenously and is easy to transfect. Endogenous proteins aggregate in discrete domains in the nucleus as shown in Fig. 30A-A'. Similar results were obtained using U2OS cells transfected with the constructs SALL1-YFP or GFP-SALL1 (Fig. 30B-B'). As a control, we used empty vectors that expressed single GFP or YFP molecules (*pEYFP-C1* or *pEGFP-N1*), which showed homogenous distribution in the nucleus and cytoplasm (Fig. 30C-C'). The expression pattern of SALL1 in U2OS was similar to the pattern previously described in the literature and we concluded that this cell line could be used as a model to study the subcellular localization of SALL1.

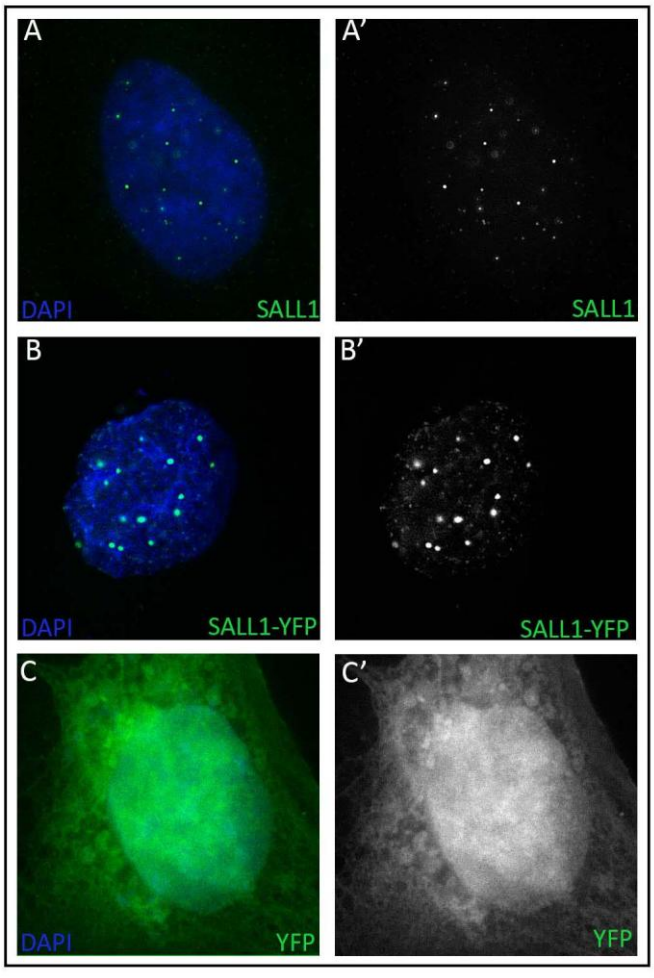


Figure 30. Localization of SALL1 in nuclear bodies in U2OS cells. Endogenous SALL1 detected with anti-SALL1 antibodies **(A)**. SALL1 fused to YFP **(B)** and YFP alone **(C)** detected with anti-GFP antibodies. SALL1 proteins and YFP are shown in green, nucleus stained with DAPI in blue. **A-C'.** Green channel is shown independently in black and white. Pictures were taken with AxioD Fluorescent microscope with 63X objective.

The mammalian cellular nucleus is a complex organelle that includes several substructures called nuclear bodies (reviewed by Dundr & Misteli 2001). Aggregation of SALL1 in nuclear foci suggested that the transcription factor could take part in the formation of one type of these nuclear bodies. In order to investigate the nature of those bodies, we performed colocalization experiments by immunofluorescence to analyze whether SALL1 is part of the most known nuclear structures such are the PML bodies or the nuclear speckles. Our results show that SALL1 does not colocalize with PML, except for a few bodies, neither with nuclear speckles as shown in Fig. 31A-B''. Therefore, we could conclude that SALL1 localizes to nuclear foci seemingly independent of PML bodies or SC-35 speckles.

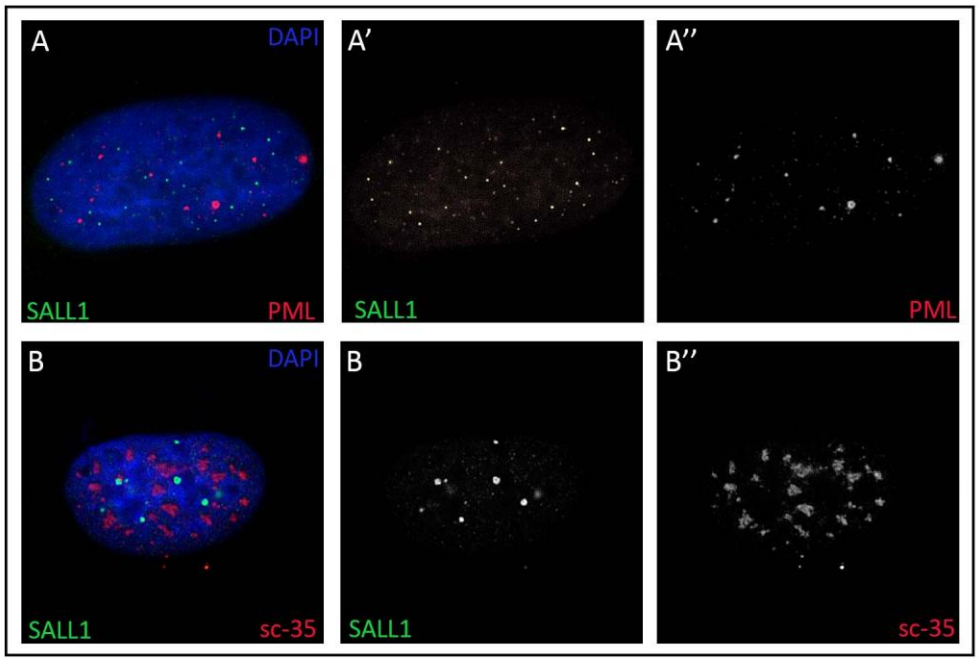


Figure 31. Comparison between SALL1 nuclear bodies and PML bodies or speckles. SALL1 endogenous protein (green) do not colocalize with PML bodies (red) **(A)** neither with speckles detected using anti-SC-35 antibodies (red) **(B)**. Nuclei were stained with DAPI (blue). **A'-B''.** Green and red channels are shown independently in black and white. Pictures were taken with an AxioD Fluorescent microscope using a 63X objective.

1.2.1. Human SALL1 partially colocalizes with SUMO proteins

Since we demonstrated the SUMOylation of human SALL1 in cultured cells, we decided to check by staining the possible colocalization between SALL1 and SUMO in mammalian cells. We transfected U2OS cells with SALL1-YFP together with bioSUMO1 or bioSUMO2 under the same conditions used in the pulldown (Fig. 32A-B''). We also used

antibodies recognizing endogenous SUMO proteins to confirm the results. As shown in Fig. 32, there was a partial colocalization between SALL1 and SUMO proteins: some of the SALL1 bodies clearly colocalized with SUMO1 and SUMO2/3, while other SALL1 bodies did not. Conversely, some SUMO1 and SUMO2/3 bodies colocalized with SALL1, while others did not (Fig. 32C-C'').

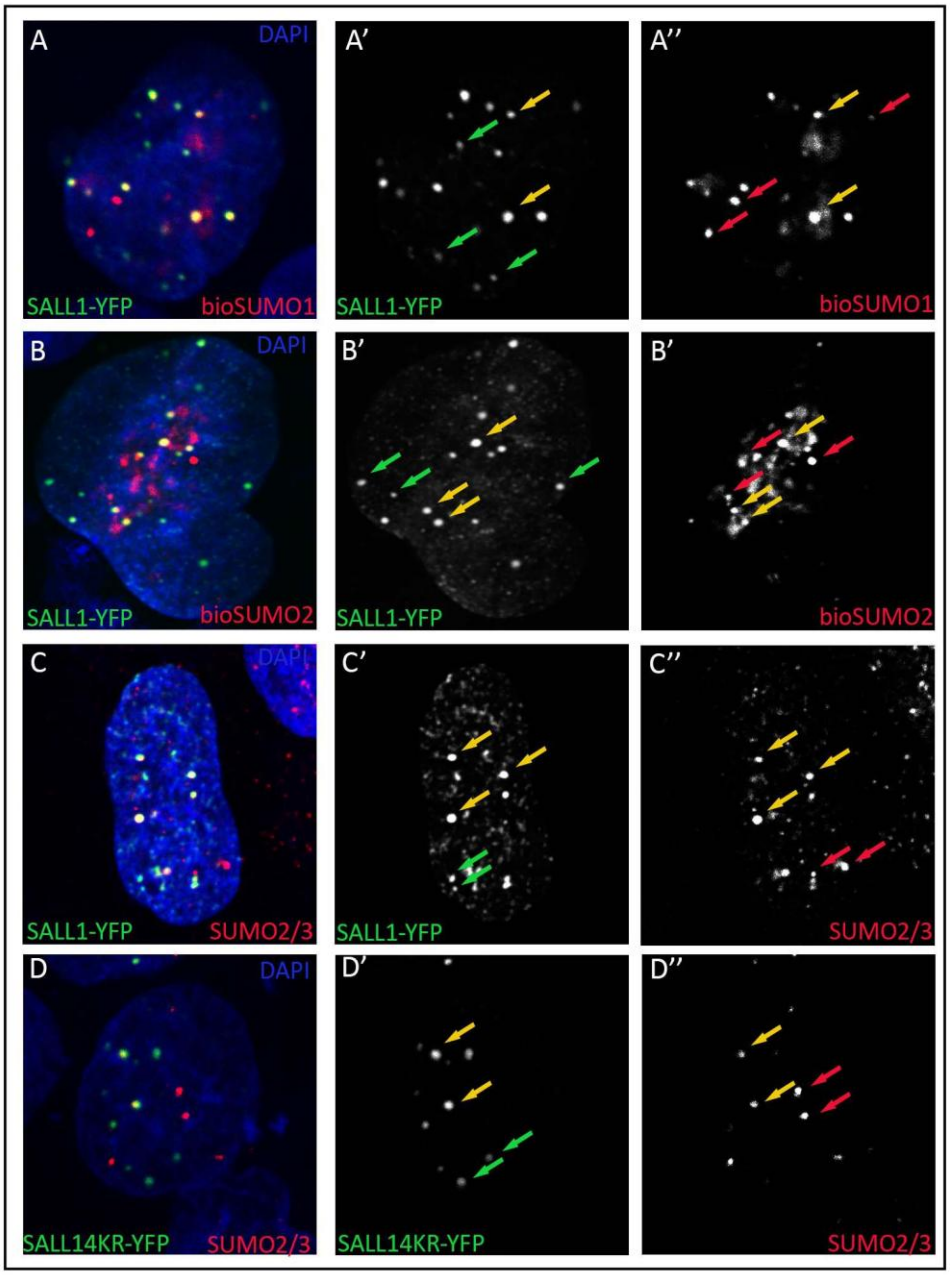


Figure 32. Partial colocalization between SALL1 proteins and SUMO in U2OS cells. SALL1-YFP (green) partially colocalizes with bioSUMO1, bioSUMO2 or endogenous SUMO2/3 (red) **(A-C).** Same result was observed for the SALL1-4KR mutant **(D)**. Yellow arrows indicate colocalization, green arrows indicate domains where only SALL1 proteins are present and red arrows indicate domains where only SUMO proteins are present. Nuclei were stained with DAPI (blue). **A'-D''** Green and red channels are shown independently in black and white. Pictures were taken with a Leica DM IRE2 confocal microscope using a 63X objective.

As described in the previous section, we showed that mutation of four lysines located in consensus sites abolished SUMOylation of human SALL1-4KR mutant. We analyzed whether the mutation of those lysines influenced the subcellular localization of the protein and, consequently, the partial colocalization between SALL1 and SUMO. First, we showed that the aggregation of SALL1-4KR in nuclear bodies is similar to that of the WT protein (Fig. 32D'). In addition, SALL1-4KR also showed partial colocalization with the SUMO proteins (Fig. 32D-D''). Taken together, our results indicated that SALL1 partial localization in SUMO bodies was not affected by its SUMOylation status.

It has been shown that PML bodies are highly SUMOylated (Zhong et al., 2000). Interestingly, we showed that not all the nuclear SUMO bodies colocalized with PML in U2OS cells (Fig. 33A-A''). Intriguingly, SALL1 and PML seemed to be mutually exclusive when localizing in most of the SUMO bodies (Fig. 5B-B'''). The possible functional meaning, if any, of this apparent mutual exclusion is unknown. This observation would deserve further analysis.

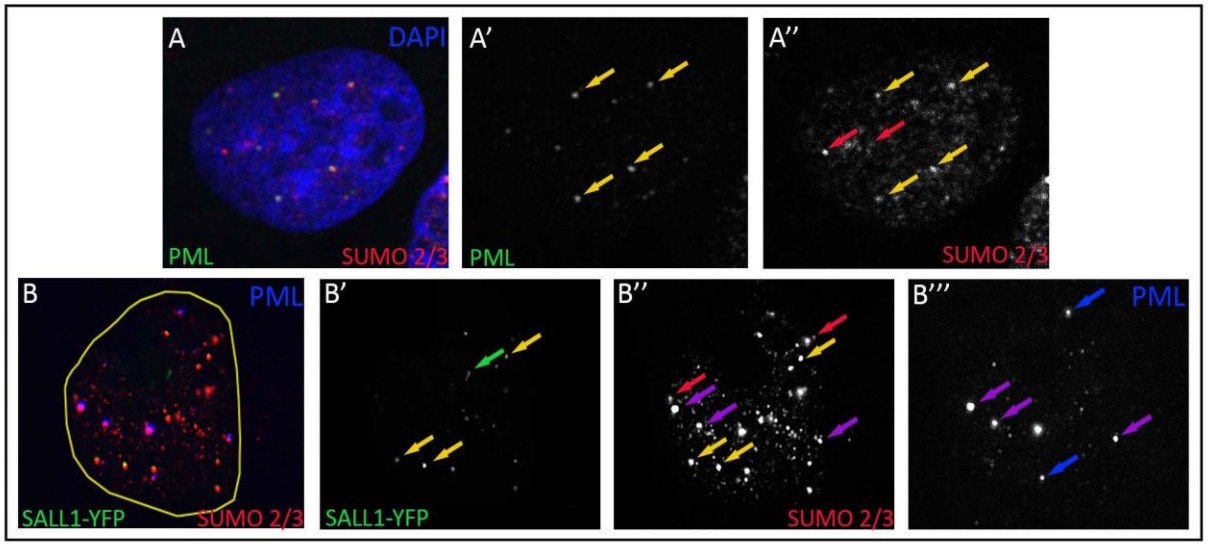


Figure 33. PML bodies in U2OS. PML bodies (green) colocalize with SUMO2/3 (red) **(A)**, but not all the SUMO spots colocalize with PML as indicated by the red arrows **(A''). B-B'''.** SALL1-YFP (green) does not colocalize with PML (blue). Yellow arrows indicate colocalization between SALL1-YFP and SUMO2/3 while violet arrows are for spots where PML and SUMO2/3 colocalize. Pictures were taken Leica DM IRE2 confocal microscope with 63X objective

1.2.2. Human SALL1 interacts with SUMO proteins by PLA

Proximity Ligation Assay (*in situ* PLA) is a technique that allows the detection of protein-protein interactions and protein modifications (Matic et al., 2010; Söderberg et al., 2006). Briefly, two primary antibodies raised in two different species are used to recognize target

antigens of interest. PLA probes, that are species-specific antibodies with a single strand DNA molecule attached, are used to bind specifically those primary antibodies. Only when the PLA probes are in close proximity generate a signal that could be detected by microscopy and indicates the presumable interaction of the antigens. The signal from each detected pair of PLA probes is visualized as an individual fluorescent spot. These PLA signals can be quantified (counted) and assigned to a specific subcellular location based on microscopy images (Söderberg et al., 2006).

We used the PLA approach to detect the subcellular interaction of SALL1 with SUMO2/3, confirming the results previously obtained. In U2OS cells, we confirmed the interaction of human SALL1-WT with SUMO2/3 in the nucleus (Fig. 34A). The vector pcDNA, used as negative control, did not show interaction with SUMO2/3, as expected (Fig. 34C). In order to analyze whether this interaction was dependent on the SUMOylation status of the protein, we analyzed the SALL1-4KR mutant. Surprisingly, we detected interaction of the mutant SALL1 with SUMO2/3, despite the fact that this mutant is not SUMOylated in cells as described in the Section 1.1 of *Results, Part II* (Fig. 34B).

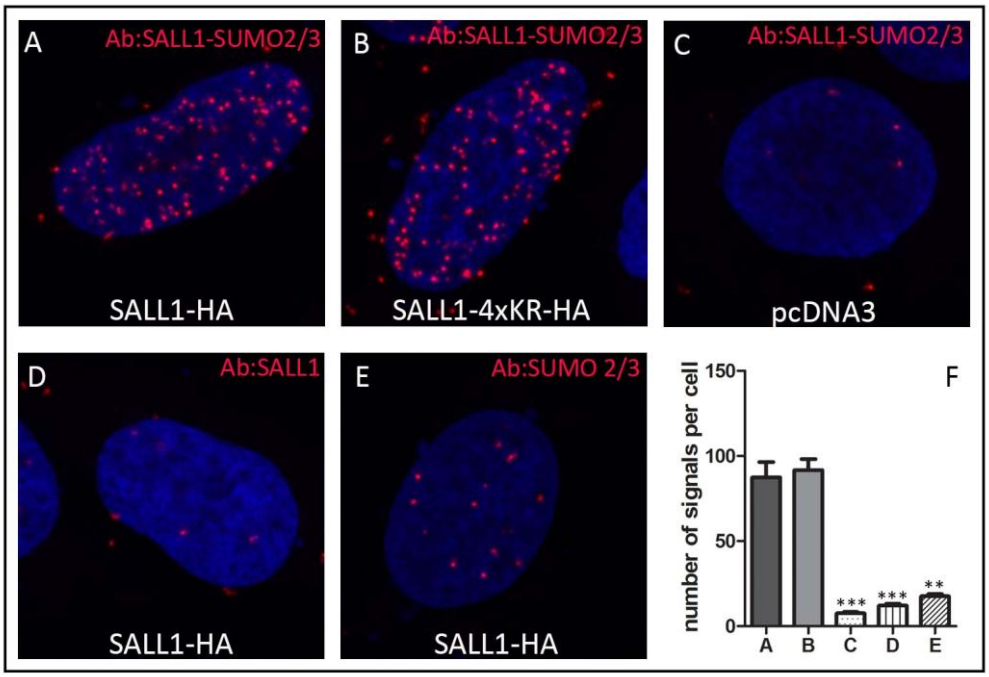


Figure 34. PLA assay to detect the interaction between SALL1 and SUMO proteins. Confocal pictures of representative PLA assays performed in U2OS cells. Red spots indicate the interaction between the two proteins. In **A** and **B,** cells were transfected with SALL1-HA and SALL1-4KR-HA, respectively. The presence of red spots indicates the interaction between SALL1 and SUMO2/3 in the nucleus. **C**. Cells were transfected with the empty vector pCDNA3 used as negative control. In **D** and **E,** cells were transfected with SALL1-HA and PLA assay was performed using separately SALL1 antibodies **(D)** or SUMO2/3 antibodies **(E)** as negative controls. **F.** Quantification of PLA signals per cell of the described conditions A, B, C, D and E. Mean difference between A and B is not significant, but both

sample shown significant difference with the control samples C, D and E as indicated by asterisks. Error bars indicate standard error of the mean. Three asterisks indicate $p<0.001$, two asterisks indicates $p<0.01$ calculated using One Way ANOVA test analysis.

As the interaction with SUMO2/3 was not dependent on the SUMOylation status of the protein, this could be due to the presence of putative SIM domains in human SALL1. To further explore this possibility, we analyzed SALL1 sequence using the GPS-SUMO software to identify putative SUMO interaction motifs. The free access software GPS-SUMO is based on a new generation group-based prediction system (GPS) algorithm integrated with Particle Swarm Optimization approach (Zhao et al. 2014; http://sumosp.biocuckoo.org). The program suggested one putative SIM between amino acids 71-75 (VLIVN) with a high probability score (45.98) and three other SIMs with lower scores: between amino acids 195-199 (VIEEN) with score 31.52, between amino acids 254-258 (ILLLA) with score 30.39 and between amino acids 1252-1256 (ISVIQ) with score 30.76. (Fig. 29C). Netzer and coworkers previously reported the direct interaction between SALL1 and SUMO1 using the yeast-two-hybrid system: the region of interaction was located between amino acids 689-1324, in the C-terminal part of the protein (Netzer et al., 2002), which is compatible with the ISVIQ site. Taken together, these results indicated the presence of at least one sequence SIM in SALL1. Further experiments would be required to demonstrate the functionality of this SIM, the possible presence of other SIMs and their putative role in SALL1 localization.

2. SUMOylation and localization of *Drosophila* Sall Proteins

Previous work from the laboratory published in 2010 and 2011 reported that *Drosophila melanogaster* Salm protein can be SUMOylated *in vitro* in the IKEE SUMOylation consensus site located between ZF2 and ZF3 (Fig. 36B) (Sánchez et al. 2010; Sánchez et al. 2011). Furthermore, SUMOylation has a role on Sall function: SUMOylation status of Sall proteins modulates the development of wing discs *in vivo*, affecting vein formation and final size of the adult wing (Sánchez et al., 2010). As the SUMOylation assays were previously done *in vitro*, we decided to check the SUMOylation capacity of Salm in *Drosophila* cultured cells using the new technology based on NeutrAvidin chromatography described in *Results, Part I*. Since Salm is not endogenously expressed in none of the available *Drosophila* cells lines, we

performed these experiments by overexpressing Salm. Kc167 and S2R+ cells were transfected with *pUAST-CFP-HisMycSalm* and with *pAc5-BirA^{opt}-T2A-GFPpuro* or *pAc5-bioSmt3-T2A-BirA^{opt}-T2A-GFPpuro* in presence of *pAc-Gal4*. *Drosophila* cells were transfected with different amounts of plasmids, treated with the proteasome inhibitor MG132 and collected at different times after transfection in order to increase the yield of Salm protein (Fig. 35). Unfortunately, cells transfected with *pUAST-CFP-HisMycSalm* died 3 days after transfection. None of our attempts yielded enough amount of *Drosophila* Salm protein to be able to perform NeutrAvidin chromatography (Fig. 35). The overexpression of Salm resulted toxic for the cells and, therefore, we decided to perform these experiments in a heterologous system, using mammalian cultured cells.

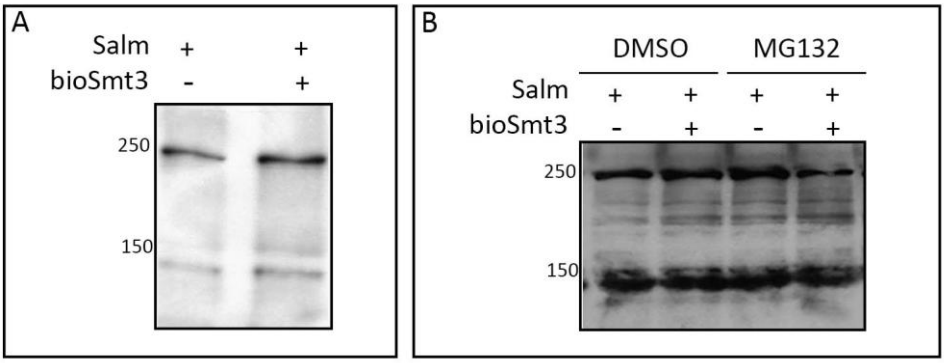


Figure 35. Overexpression of Salm in S2R+ cells. A. Overexpression of Salm in presence (+) or absence (-) of bioSmt3. **B.** Overexpression of Salm in presence or absence of MG132 with or without bioSmt3.

2.1. SUMOylation of *Drosophila* Salm in mammalian cells

To avoid the yield limitations faced with the use of *Drosophila* cells, we used HEK 293FT cells as a system to investigate the SUMOylation of *Drosophila* Salm protein. First, we checked the level of protein expression after transfecting the cells with *CMV-CFP-HisMycSalm* or with *CMV-CFP-HisMycSalm^{IKDP}*, a Salm SUMOylation mutant where the IKEE (Lys 641) consensus site was mutated to IKDP (Fig. 36B; (Sánchez et al., 2010). Once we obtained optimal levels of expression (Fig. 36A, Input panel), NeutrAvidin pulldowns were performed. Our results showed that *Drosophila* Salm WT was modified by bioSUMO1 and bioSUMO3 in mammalian cells, but that SUMOylation was abolished when the Salm^{IKDP} mutant was expressed (Fig. 36A, Elution panel). These experiments confirmed the results previously obtained *in vitro* (Sánchez et

al., 2010) and surprisingly showed that, also in the cellular environment, mutation of a single SUMOylation consensus site is enough to abolish Salm SUMOylation.

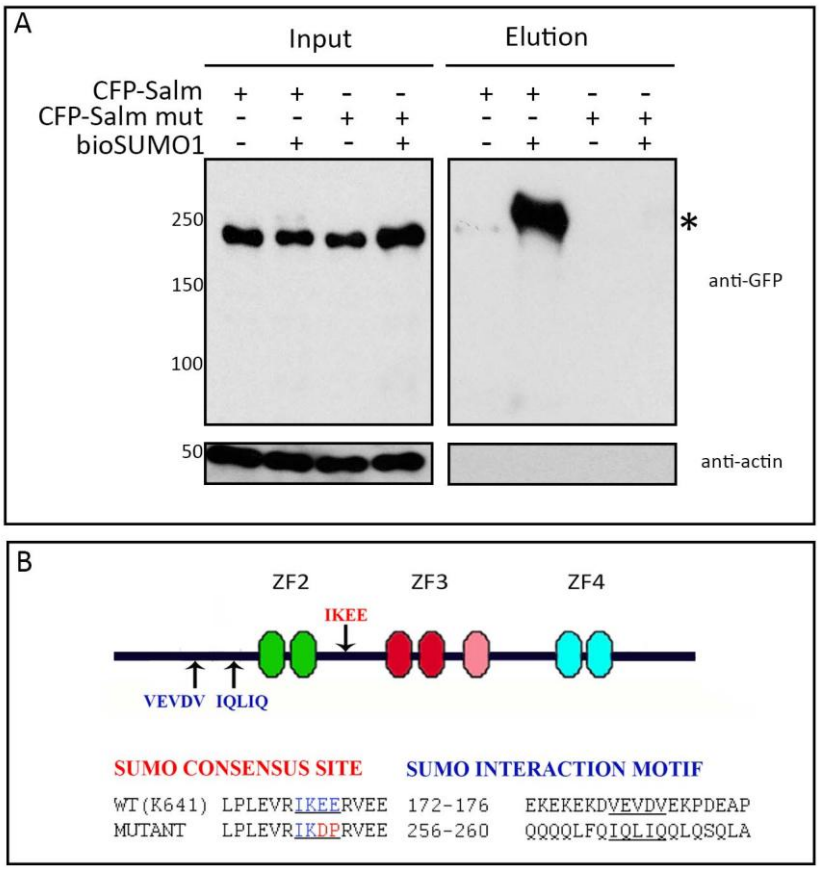


Figure 36. SUMOylation of Salm. A. Western blot of pulldowns to check SUMOylation of Salm in HEK 293FT cells. Salm fused to CFP is SUMOylated in presence (+) of bioSUMO1 as indicated by the asterisk. SUMOylation occurs in IKEE site and the mutation IKPD abolishes Salm modification (Elution, lane 4). Anti-actin antibodies were used as a loading control. Molecular weight markers are shown to the left. **B.** Salm schematic representation. In red, the SUMO consensus site mutated in CFP-Salm mutant and in blue the predicted SIMs are shown.

2.2. *Drosophila* Salm localizes in nuclear bodies in cultured cells

Despite the low expression of the *pUAST-CFP-HisMycSalm* construct in *Drosophila* cells, it was still possible to observe the localization of Salm in the transfected cells. *pUAST-CFP-HisMycSalm* plasmid was used to transfect S2R+ *Drosophila* cells in presence of *pAc-Gal4*. The fusion protein localized in discrete nuclear domains, as it was shown *in vivo* in the PG (See below Fig. 37). Furthermore, partial colocalization between Salm and endogenous Smt3 was demonstrated (Fig. 37A-A'').

Since we proved the SUMOylation of *Drosophila* Salm in mammalian cells, we decide to check, as we did with the human homolog SALL1, its colocalization with SUMO proteins in human cells. U2OS cells were used to perform immunofluorescence experiments. As it can be shown in Fig. 37B-B'', partial colocalization existed between Salm and endogenous SUMO proteins. As described in the previous section, we showed that a mutation of the consensus site abolishes SUMOylation of *Drosophila* Salm^{IKPD}. We analyzed whether the mutation of IKEE consensus site influenced the subcellular localization of the protein and, consequently, the partial colocalization between Salm and SUMO. First, we showed that, as previously shown, the aggregation of Salm^{IKPD} in nuclear bodies was similar to that of the WT protein (Fig. 37C; Sánchez et al. 2010). In addition, Salm^{IKPD} still showed partial colocalization with SUMO2/3 (Fig. 37C-C''). Taken together, our results indicate that Salm partial localization in SUMO bodies is not affected by its SUMOylation status in U2OS mammalian cells.

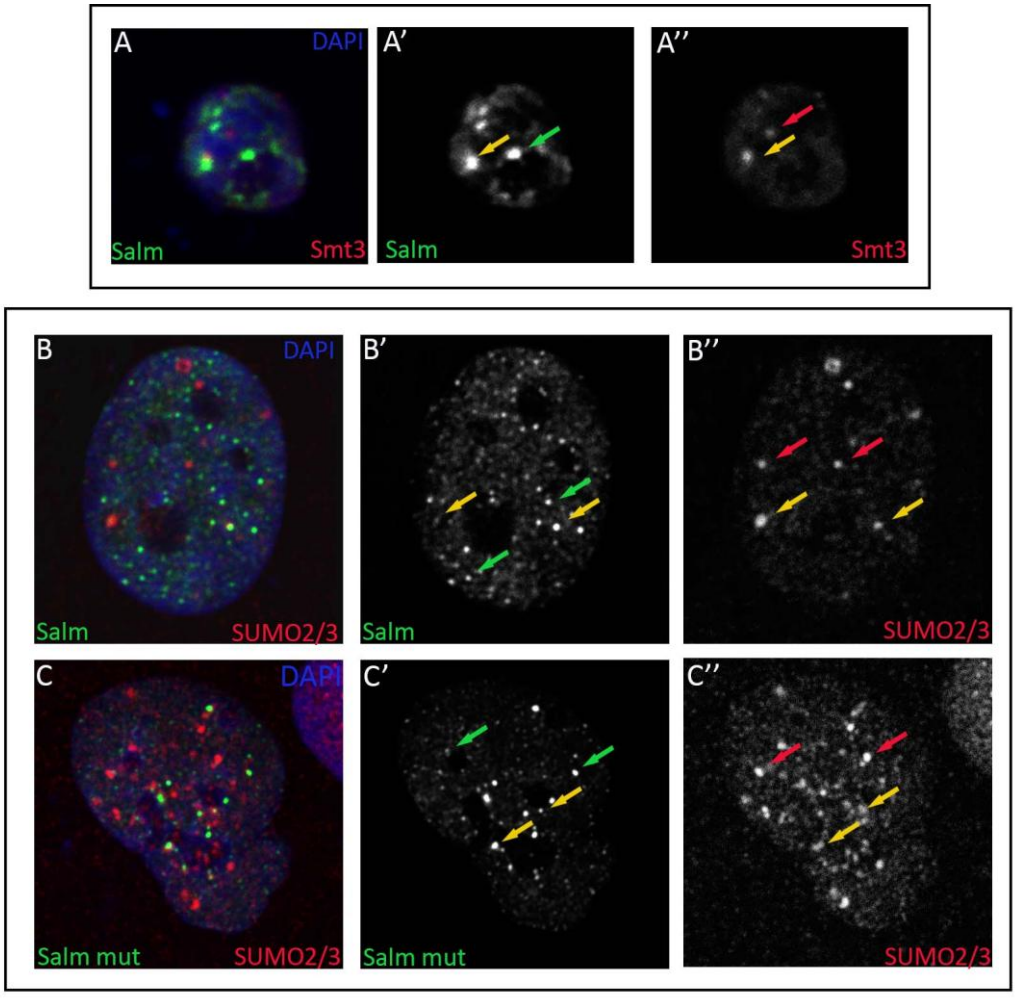


Figure 37. Partial colocalization between Salm and SUMO. Salm-CFP (green) partially colocalized with Smt3 (red) in S2R+ *D. melanogaster* cells **(A-A'').** In **B-B'',** the partial colocalization between Salm and SUMO2/3 in U2OS cells is shown. Yellow arrows indicate colocalization, green arrows indicate

domains where only Salm is present and red arrows indicate domains where only SUMO2/3 is present. Similar results were observed for Salm mutant **(C-C'').** Nuclei were stained with DAPI (blue). Green and red channels are shown independently in black and white **(A'-C'')**. Pictures were taken using a Leica DM IRE2 confocal microscope with a 63X objective.

These results are similar to what we observed for human SALL1, where the interaction with SUMO2/3 is not dependent on the SUMOylation status of the protein and could be due to the presence of SIM domains in the sequence. Also in this case, we analyzed the *Drosophila* Salm sequence using the GPS-SUMO software and the program suggested two putative SIMs: the first between amino acids 256-260 (IQLIQ) with score 24.94 and another between amino acids 172-176 (VEVDV) with lower probability score (19.35) (Fig. 36B). Further experiments would be required to demonstrate the presence of this SIM and its putative function in Salm localization.

3. Analysis of CBX4 as a possible SUMO E3 ligase for SALL1

Our results demonstrated the SUMOylation of human SALL1 in HEK 293FT cells. As SUMOylation might be regulated *in vivo* by E3 ligases, we were interested on studying which ligases could be involved in SALL1 SUMOylation and analyzing their possible interactions *in vivo*. Therefore, our aim of this part of the work was to identify possible E3 ligases for SALL1 using different approaches.

3.1. Identification of human SALL1 partners by BioID

The "proximity proteomics" BioID methodology in combination with Mass Spectrometry was used to identify possible interactors of human SALL1. BioID is a method to screen for proximate and interacting proteins in living cells. The method is based on a modified form of the biotin ligase, called BirA*, which losses target specificity and is able to add biotin to unspecific sequences in close proximity. BirA* is fused to a protein of interest and expressed in cell culture, where the efficient biotinylation of endogenous proteins adjacent to, or interacting

with, our protein of interest takes place by adding an excess of biotin in the culture medium (Fig. 38A) (Roux et al. 2012; Dingar et al. 2014).

Myc-BirA* was fused to the N-terminus part of human SALL1, either the full length form (FL) or the truncated form SALL1(826) present in Townes-Brocks patients. Lysates of HEK 293FT cells transfected with *CMV-Myc-BirA*-SALL1* or *CMV-Myc-BirA*-SALL1(826)* were used to perform pulldown experiments using NeutrAvidin beads. Total protein eluates were resolved on a SDS-PAGE and stained with Colloidal Coomassie (Fig. 38B). Bands were excised from the gel, subjected to trypsin digestion and, analyzed by Mass Spectrometry. The experiments were done in triplicates. In collaboration with the group of Dr. Jesper Olsen at the University of Copenhagen, we identified proteins that are possible interactors of human SALL1, either the FL or the truncated forms. We identified 2608 proteins from this Mass Spectrometry analysis. After our analysis of the ratio of intensities between the SALL1(FL) and the SALL1(826) samples (explained in *Materials and Methods*), we obtained a list of 404 possible interactors for SALL1(FL). We focused our attention in those proteins, the other proteins interacting with the truncated SALL1(826) will be the object of a different study in the laboratory (Appendix V).

We performed a bioinformatics analysis of the selected proteins (Fig. 39). GO analysis revealed that the putative interactors of SALL1 were involved in different biological processes as translation, catabolic process of transcribed mRNA or protein targeting to membrane. The 404 proteins were localized in different cellular compartments: in the nucleus, they were related to the SWI/SNF complex, PML body or as proteins associated to chromatin; The portion of proteins present in the cytoplasm were associated to the ribosome, extracellular vesicular exosome and collagen. In respect to the molecular function, most of the proteins identified were factors involved in gene expression regulation. As shown in Fig. 39, GO categories emerging from the analysis were phosphoproteins, chromatin, mRNA, DNA or histone deacetylase binding proteins, as well as proteins involved in DNA catalytic activity. Interestingly, the SUMO ligase activity is well represented. Among the interactors of SALL1, we identified CBX4, PIAS1, PIAS2 and PIAS4, known mammalian SUMO E3 ligases which could be involved in the SUMOylation of SALL1 *in vivo*.

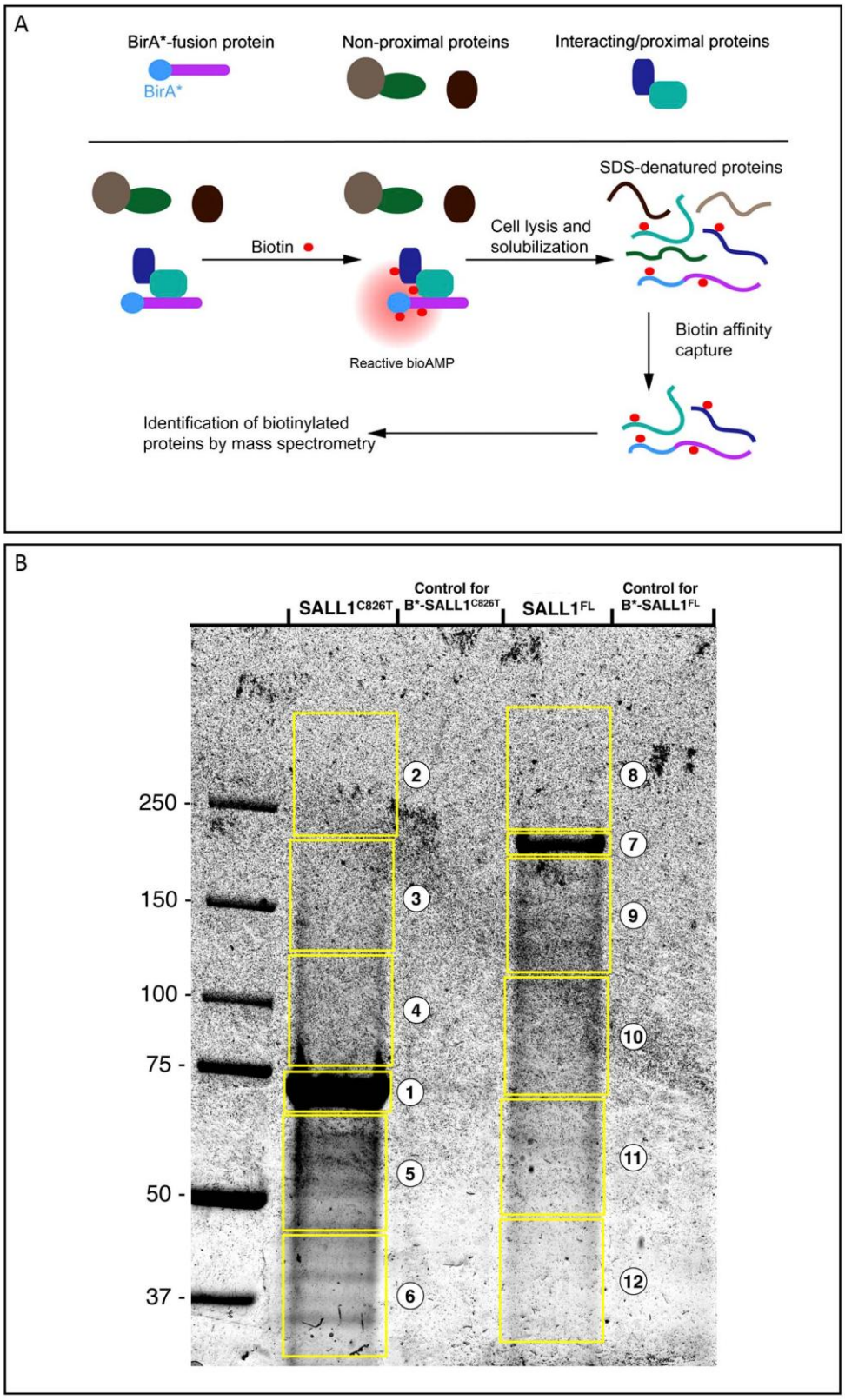


Figure 38. The proximity proteomics BioID methodology in combination with Mass Spectrometry. A. Schematic representation of the BioID methodology adpted from Roux et al. 2012. **B.** Gel stained with Colloidal Coomassie. Pulldowns of HEK 293FT cells transfected with *CMV-Myc-BirA*-SALL1(826)* or *CMV-Myc-BirA*-SALL1(FL)* with their respective controls. Molecular weight markers are shown to the left. Excised gel fragments used for Mass Spectrometry are numbered.

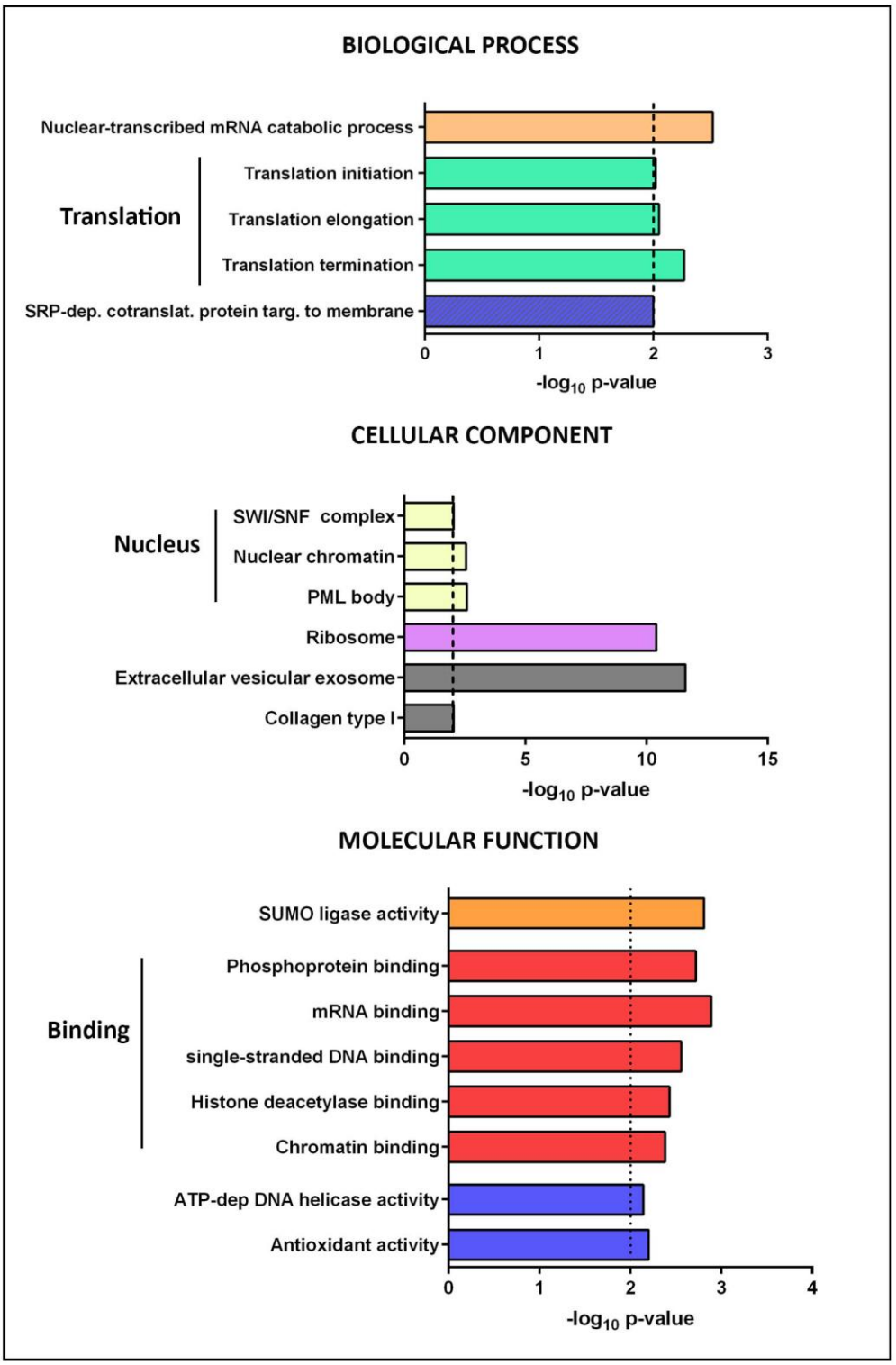


Figure 39. Analysis of SALL1(FL) interactors by Mass Spectrometry. GO analysis for biological process, cellular component and molecular function performed using Innate DB.

CBX4 is the homolog of *Drosophila* Pc. Interestingly, the genetic interaction between *sall* genes and *Pc* was previously reported in *Drosophila*, as mutations in *salm* enhanced

Pc group mutations phenotype during embryogenesis (Casanova 1989; Landecker et al. 1994). Therefore, we decided to analyze the possible function of CBX4/Pc as SALL E3 ligase.

3.1.1. Validation of the interaction of SALL1 and CBX4

To validate the Mass Spectrometry data we used Western blot as a complementary technique. BirA* was fused to the C-Terminus of CBX4 in a vector containing the CMV promoter. *CMV-CBX4-BirA** was used to transfect HEK 293FT cells in combination with *CMV-SALL1-2xHA* or *CMV-SALL1(826)-2xHA*, where the FL or the truncated forms of SALL1, respectively, were fused to two copies in tandem of the hemagglutinin antigen HA. As negative controls, cells transfected separately with the individual plasmids were used. After transfection, cell lysates were subjected to pulldown using NeutrAvidin beads and the eluates were analyzed by Western blot using anti-HA-antibodies (Fig. 40). CBX4 appeared as interactor of both proteins, the FL and the truncated SALL1(826), as is shown in the Elution panel (Fig. 40). Each lane of the gel was loaded with the same amount of total proteins, which was confirmed by using anti-GAPDH antibodies as control, but the relative amount of SALL1(FL) and SALL1(826) was different as is clearly visible in the panel on Input. Despite the presence of more exogenous SALL1(826) in cells, CBX4 seemed to bind preferentially to SALL1(FL). These experiments confirmed the results from Mass Spectrometry described above.

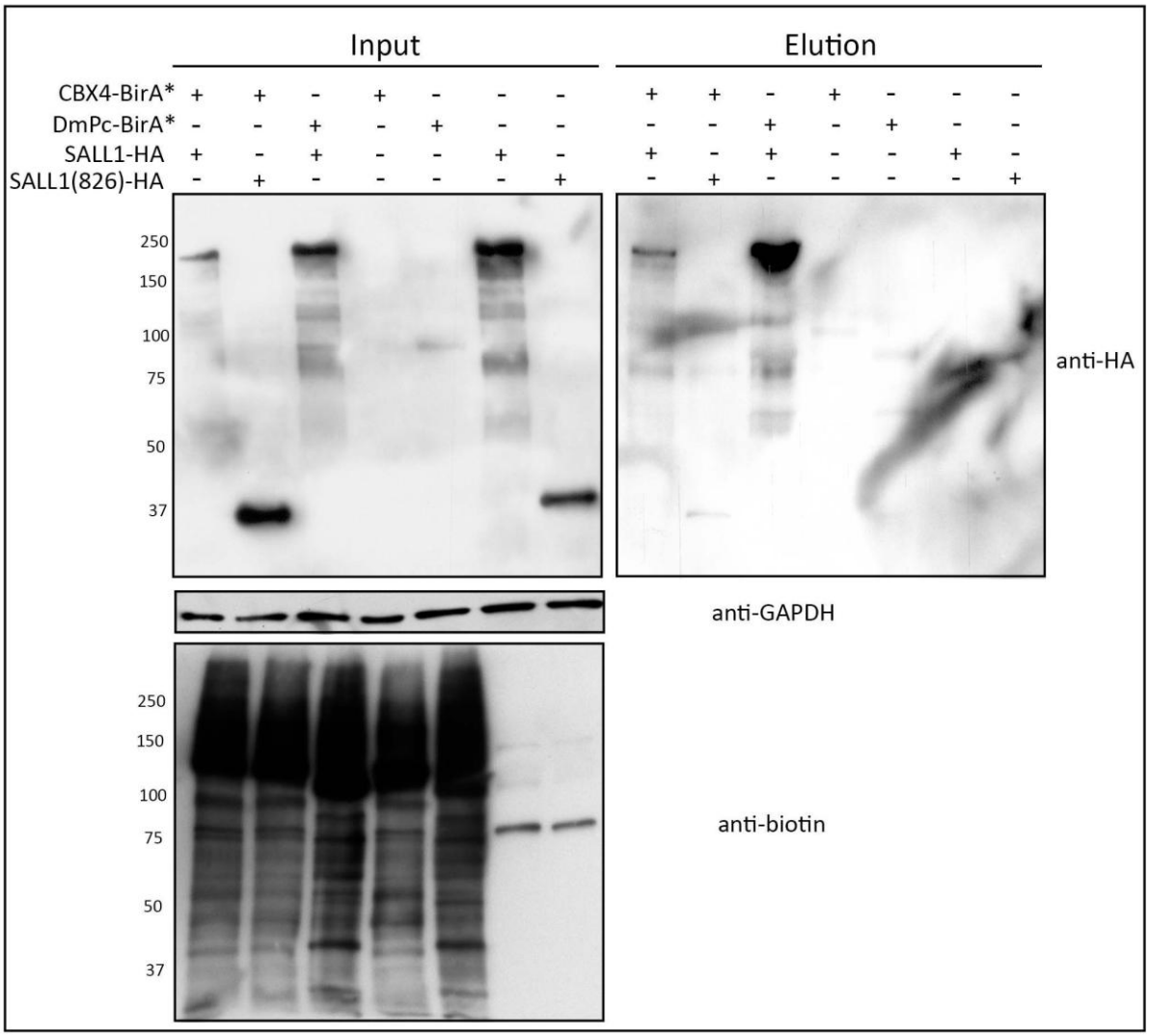


Figure 40. Validation of the binding of CBX4/Pc with SALL1. In the Input panel, the relative expression of the SALL1 proteins fused to HA tag is shown. In the elution, SALL1(FL)-HA binds preferentially to CBX4-BirA* (lane1) and to DmPc-BirA* (lane 3). Controls are shown in lanes 4-7. Anti-GAPDH and anti-biotin antibodies were used as loading control and to show the efficiency of the different pulldowns, respectively. Molecular weight markers are shown to the left.

3.2. SALL1 and CBX4 interact directly

Our previous results confirmed the interaction between SALL1 and the E3 ligase CBX4. However, due to the nature of the BioID method, this interaction could be direct or indirect, being mediated by a third protein. In order to discriminate between these two possibilities, we expressed SALL1 and CBX4 in cultured cells and performed copulldown experiments. HEK 293FT cells were transfected using *CMV-SALL1-GFP* and *CB6-HA-CBX4* plasmids. Two days after transfection, cells were lysated and incubated with anti-GFP beads

(called GFP-trap; see *Materials and Methods*) to allow the capture of the SALL1-GFP fusion protein. After elution, the material was analyzed by Western blot using anti-HA antibodies. As it can be seen in Fig. 41, CBX4 was present in the Elution panel of the pulldown, indicating its direct interaction with SALL1-GFP. Those results confirmed the data previously obtained with the BioID approach and indicated that the interaction between the two proteins is direct and does not require a third party.

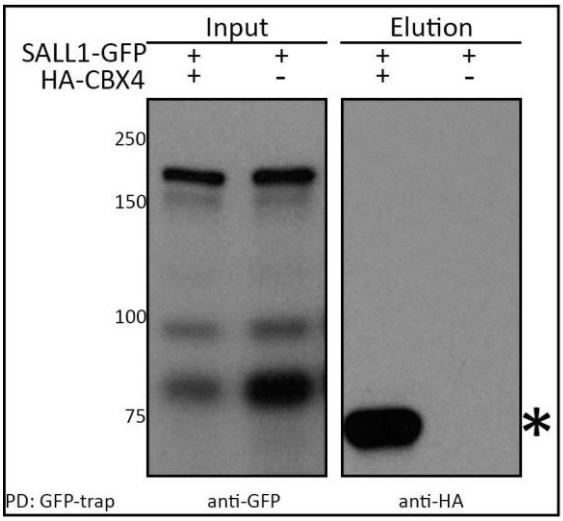


Figure 41. Direct interaction of SALL1 and CBX4 by co-pulldown. In the Input panel, SALL1 is shown by using anti-GFP antibodies in presence (+) or absence (-) of CBX4. In the Elution panel, anti-HA antibodies were used. Molecular weight markers are shown to the left. Asterisk indicates the HA positive band.

3.2.1. CBX4 localizes in nuclear bodies not overlapping with SALL1

PcG proteins are concentrated in nuclear foci called PcG bodies (reviewed by Pirrotta & Li 2012). In order to detect the localization of CBX4, we contributed to the characterization of the new anti-CBX4 antibodies developed by the company Proteintech. We tested the specificity of the antibodies by Western blot in lysates from mammalian cells, where we showed a major band that appeared at 75 KDa (isoform A) and a smaller band around 50 KDa that corresponded to the isoform B (Fig. 42A). In addition, we tested the localization of CBX4 in U2OS cells by immunofluorescence, confirming the presence of CBX4 in nuclear structures similar to the PcG bodies (Fig. 42B-B'').

Since the interaction between SALL1 and CBX4 was demonstrated, we decided to test whether these proteins colocalized in the cells. *CMV-SALL1-YFP* plasmid was transfected into U2OS cells, where endogenous CBX4 was visualized by immunofluorescence using the anti-CBX4 antibodies. Surprisingly, our confocal pictures showed that there was not colocalization of SALL1 and CBX4 in the nuclear bodies (Fig. 42C-C''). In view of these results, we thought that this interaction could take place, not in the nuclear bodies, but in the nucleoplasm. In order to check that, we decided to apply PLA to confirm our hypothesis.

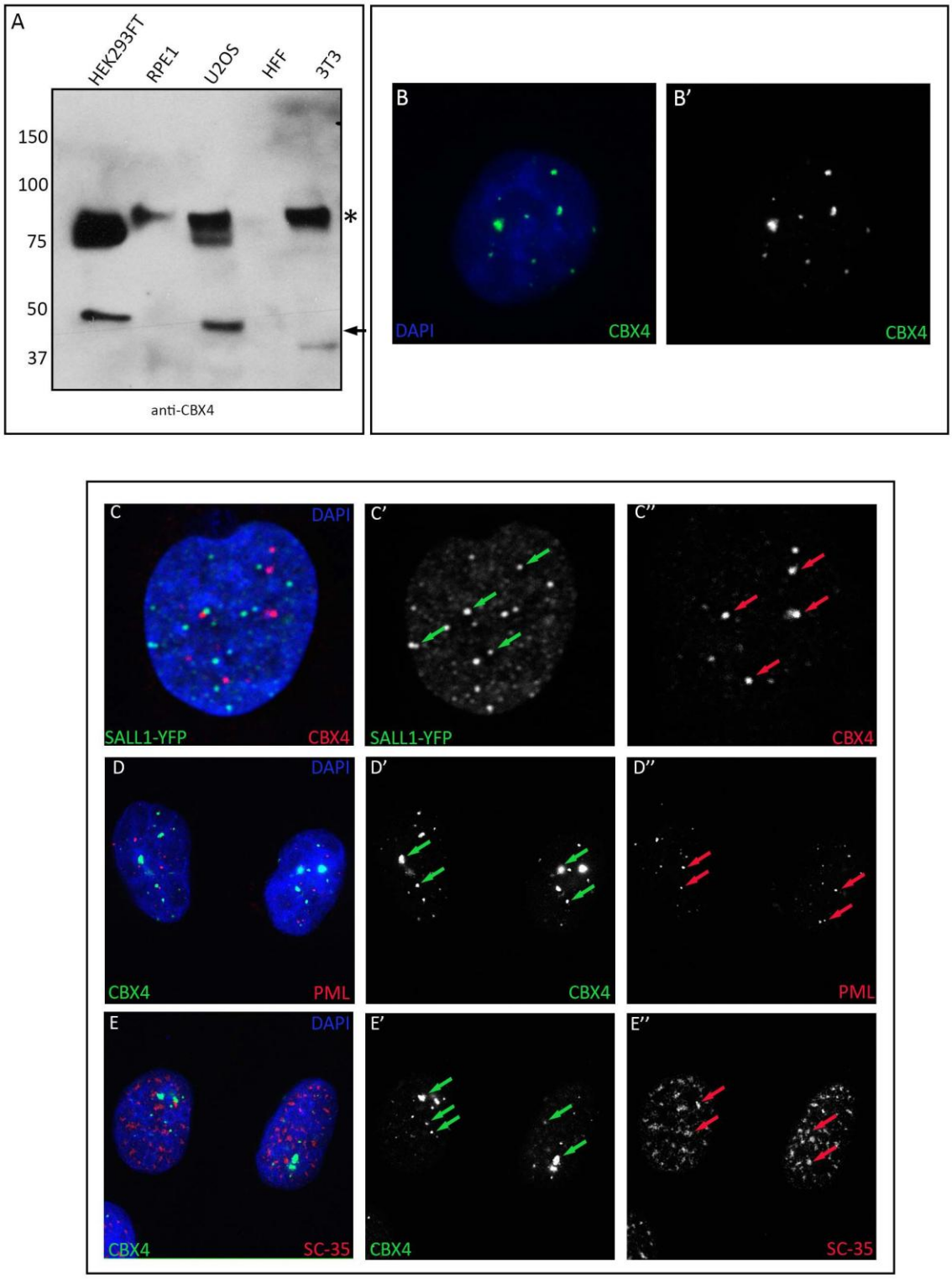


Figure 42. Characterization of the anti-CBX4 antibodies (Proteintech). A. Western blot using extracts from different cell lines. The major band corresponding to the isoform A of CBX4 was identified at 75 KDa, as expected (asterisk). The isoform B is indicated by an arrow. Molecular weight markers are shown to the left. **B-B'.** CBX4 localization in U2OS cells detected by the anti-CBX4 antibody using an AxioD fluorescent microscope. **C-C''.** Confocal pictures of U2OS cells transfected with SALL1-YFP. SALL1 domains (green) and Pc bodies (red) do not colocalize. CBX4 (green) did not colocalize neither with the PML bodies (red) **(D-D''),** nor with speckles **(E-E'')** visualized using anti-SC-35 antibodies (red). B'-E''. Green and red channels are shown independently in black and white. Nuclei were stained with DAPI (blue). Pictures in **C-C''** were taken using a Leica DM IRE2 confocal microscope with a 63X objective. Pictures **D-E''** were taken using a Leica DM IRE2 confocal microscope with a 40X objective.

U2OS cells were transfected with *CMV-SALL1-YFP* and anti-SALL1 and anti-CBX4 antibodies were used to perform PLA. Our analysis of the number of spots confirmed the interaction between SALL1 and CBX4 in the nucleus, when compared to cells transfected with the empty vector *pcDNA* used as negative control (Fig. 43A, B). PLA analysis was done taking into account only the cells transfected, which were YFP positive. Controls to test the background of different antibodies were done (Fig. 43C, D). It was not possible to use anti-GFP antibodies because they generated nonspecific positive spots. In conclusion, our experiments showed the direct interaction of SALL1 and the E3 ligase CBX4 in the nucleus, probably in the nucleoplasm, given the absence of colocalization in the nuclear bodies.

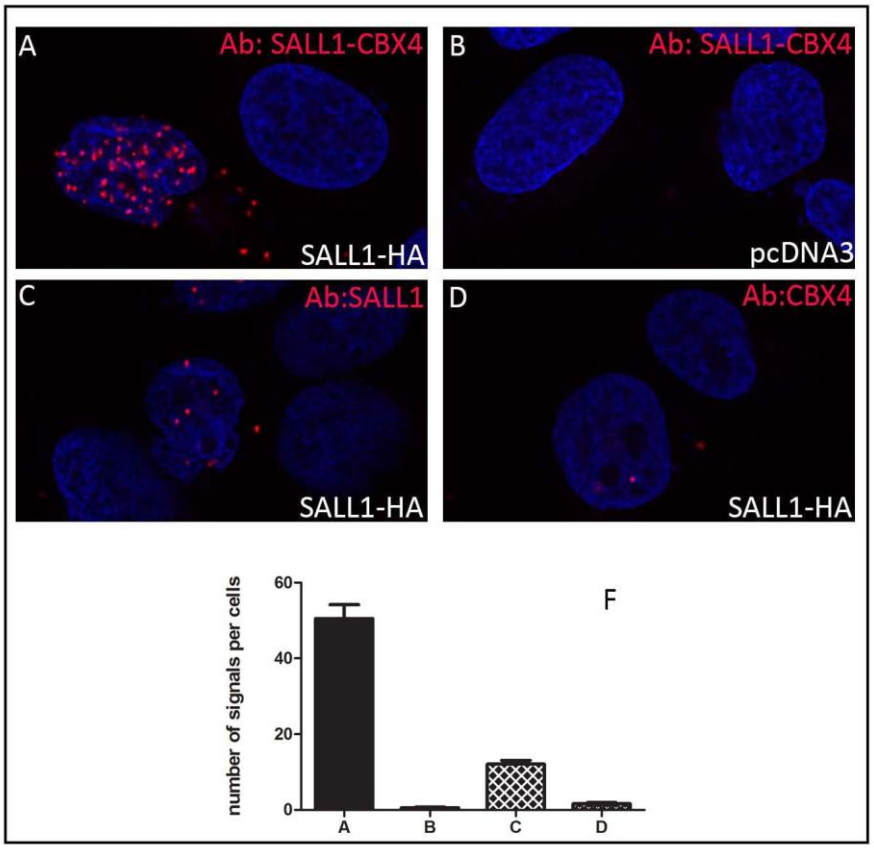


Figure 43. PLA assay for SALL1 and CBX4 colocalization. A-D. Confocal pictures of PLA assays performed in U2OS cells. Red spots indicate the interaction between the two proteins. **A.** Cells were transfected with SALL1-HA. The presence of red spots indicates the nuclear interaction between SALL1 and CBX4. **B.** Cells were transfected with the pCDNA3 empty vector as negative control. **C, D.** Cells were transfected with SALL1-HA and PLA assays were performed using only SALL1 antibodies **(C)** or CBX4 antibodies **(E),** separately, as negative controls. **D.** Quantification of PLA signal per cell for the described conditions A, B, C and D. Mean of sample A shows significant differences with mean of control samples B, C and D as indicated by asterisks. Error bars indicate standard error of the mean. Three asterisks indicate p<0.001calculated using One Way ANOVA test analysis.

In addition to the colocalization with SALL1, we analyzed by immunofluorescence the colocalization of CBX4 with other nuclear factors, such as PML or SC-35 (Fig. 42D-E'').

We did not observe colocalization of CBX4 neither in PML bodies, nor in speckles, suggesting that there was not interaction with these factors in the nuclear bodies. However, we cannot discard that the interaction takes place in the nucleoplasm, as we did not proceed with further experiments such is the PLA analysis.

3.3. Functional analysis of SALL1-CBX4 interaction

Our PLA experiments proved the interaction between SALL1 and CBX4 in the nucleoplasm. We hypothesized that this interaction might influence the SUMOylation status of SALL1 in the cells, based on the E3 ligase capacity of CBX4.

3.3.1. CBX4 influences the SUMOylation status of SALL1

As mentioned in the *Introduction*, the presence of exogenous CBX4 in mammalian cells enhances the SUMOylation of CTBP1 (Kagey et al., 2003). In order to analyze the possible effect of CBX4 on SALL1 SUMOylation, we transfected HEK 293FT cells with *CMV-SALL1-YFP*, together with *CAG-bioSUMO1-T2A-BirA^{opt}-T2A-GFPpuro* or *CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro* in presence or absence of *CB6-HA-CBX4*. Two days after transfection, cells were collected and lysed and pulldowns were performed using the NeutrAvidin beads. Our results confirmed the SUMOylation of SALL1 in presence of bioSUMO1 (Fig. 44A). Surprisingly, when CBX4 was added to the cells, the SUMOylated form of SALL1 disappeared, as it is shown in Fig. 44B, where SALL1-YFP is visible in the Input but it is not detectable in the Elution. Using anti-HA antibodies we revealed a strong SUMOylation of CBX4 in the cells (Fig. 44C). Therefore, we hypothesized that the bioSUMO1 or 3 (data not shown) present in the cells were massively incorporated into CBX4, limiting the availability of the SUMOylation machinery to be conjugated to SALL1-YFP.

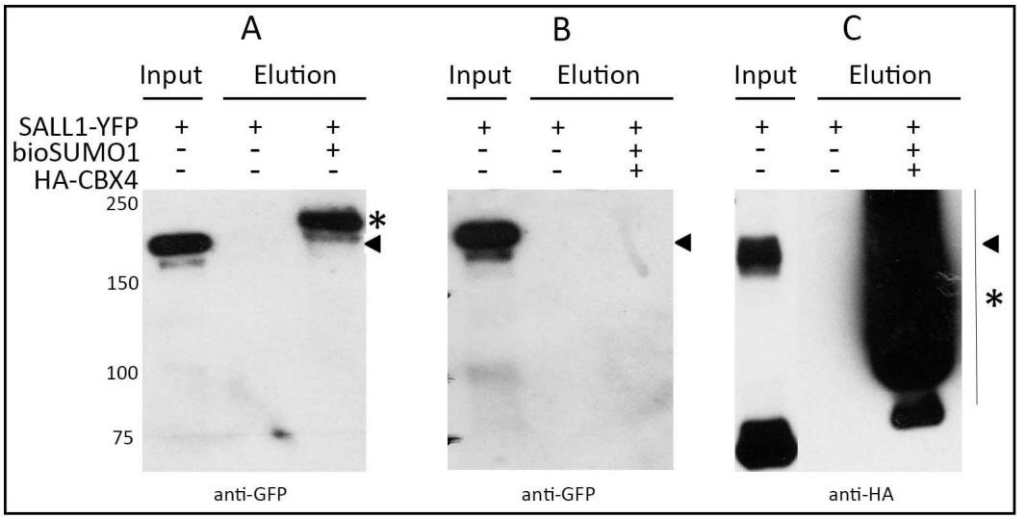


Figure 44. SUMOylation assay for SALL1 in presence of CBX4. A. SUMOylation of SALL1-YFP, marked by an asterisk, is shown in presence of bioSUMO1. **B.** The presence of CBX4 (+) impedes the SUMOylation of SALL1 (Elution, lane 2). **C.** Strong SUMOylation of HA-CBX4 in presence of bioSUMO1 (Elution, lane 2) indicated by the line and the asterisk. Arrowheads mark the unmodified SALL1. Molecular weight markers are shown to the left.

As our experiments of SUMOylation in cells were inconclusive, we decided to take an alternative approach. In collaboration with Dr. Manuel S. Rodriguez and Dr. Valerie Lang (Inbiomed, San Sebastian) i*n vitro* SUMOylation assays were performed. The results showed that the SUMOylated form of SALL1 FL in presence of SUMO1 or SUMO2/3 varied when different amounts of CBX4 were added to the reaction (Fig. 45). The absence of SUMOs and CBX4 were used as negative controls. However, we experienced different technical problems during the performance of these experiments, mainly difficulties to express SALL1 in sufficient quantities for the experiments due to its large size. In summary, our *in vitro* results are compatible with a role of CBX4 promoting the SUMOylation of SALL1, although the experiments would need to be confirmed.

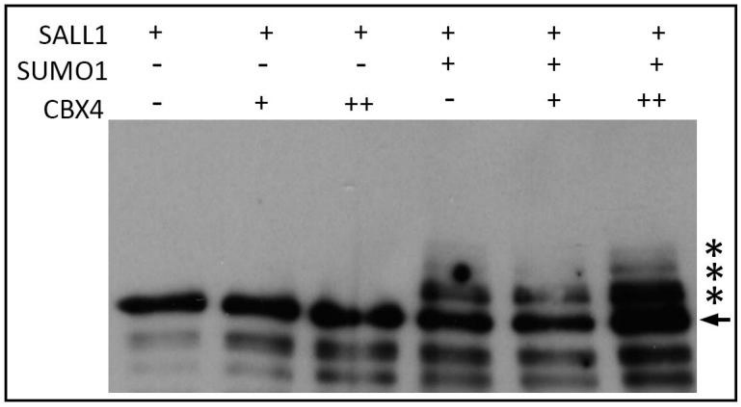


Figure 45. In vitro SUMOylation of SALL1 in presence of CBX4. Presence of CBX4 at two different quantities (+ and ++) contributed to the SUMOylation of SALL1 in presence of SUMO1 (lane 5 and 6). Asterisks indicate the SUMOylated forms of SALL1. Arrow indicates the unmodified SALL1.

3.3.2. SALL1 influences the number of PcG bodies

During the course of the SUMOylation experiments in cultured cells described in the previous section, we observed that the number of PcG bodies detected with the CBX4 antibodies varied in presence of SALL1. Therefore, we hypothesized that SALL1 has a role on PcG bodies formation. In order to prove this, we transfected U2OS cells with *CMV-SALL1-YFP*, *CMV-SALL1(826)-YFP* or *CMV-EGFP-β-galactosidase*, used as a negative control. PcG bodies were detected by immunofluorescence using anti-CBX4 antibodies. Pictures were taken with an automated fluorescent microscope and were analyzed using the MetaExpress software. With the help of Dr. M. González in the laboratory, we set up a specific module that count the number of PcG bodies specifically in the transfected cells. Our results showed that the presence of SALL1-YFP FL increased significantly the number of PcG bodies in a 27% with respect to the control (Fig. 46A). The expression of the truncated form SALL1(826)-YFP also increased the number of PcG bodies in a 13% with respect to the control, but this difference was not significant.

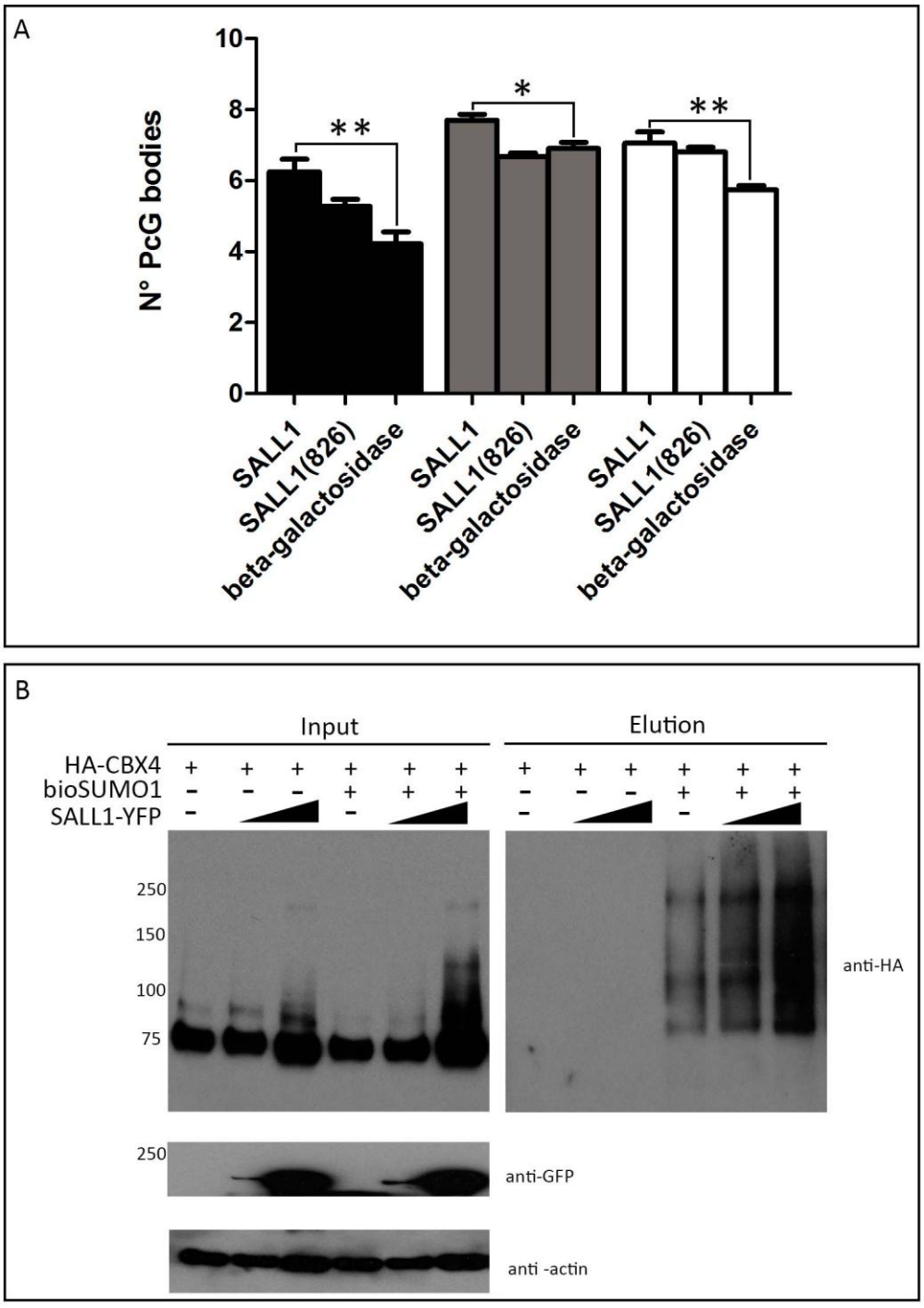


Figure 46. SALL1 influences the levels of CBX4. A. Representation of 3 independent experiment performed in U2OS transfected with 500 μg of SALL1-YFP, SALL1(826)-YFP or β-galactosidase-GFP used as negative control. Number of Pc bodies increased significantly in presence of SALL1-YFP. One asterisk indicates $p< 0.05$, two asterisks $p<0.01$, calculated using Student's *t* test analysis. Error bars indicate standard deviation. **B.** Western blot showing that, in presence of SALL1, the amount of CBX4 increased. This effect is visible in the Input (lanes 3 and 6) and consequently in the Elution (lane 6). Anti-GFP antibodies were used to control the relative amount of SALL1 added into the cells and anti-actin antibodies were used as loading control. Molecular weight markers are shown to the left.

We used Western blot analysis as a complementary technique to analyze whether the increase in number of PcG bodies was due to increased levels of CBX4. HEK 293FT cells were transfected with *CB6-HA-CBX4*, *CAG-bioSUMO1-T2A-BirA^{opt}-T2A-GFPpuro* or *CAG-bioSUMO3-T2A-BirA^{opt}-T2A-GFPpuro*, and different amounts of *CMV-SALL1-YFP*. We found a direct correlation between the amount of CBX4 visible in each lane and the amount of SALL1-YFP added to the cells (Fig. 46B, Input panel). We detected also an increased amount of SUMOylated CBX4 in presence of SALL1 (Fig. 46B, Elution panel). However, the increased CBX4 levels were detected also in absence of bioSUMO1, indicating that the correlation between SALL1 and CBX4 levels might not depend on the SUMOylation status of these proteins. The increase in SUMOylated CBX4 partially reflected the higher levels of CBX4 in the input of the pulldown reactions.

Taken together, our results suggested a reciprocal effect between SALL1 and CBX4: on one hand, CBX4 could contribute to SALL1 SUMOylation while, on the other hand, SALL1 could influence the levels of CBX4 and the formation of PcG bodies. The functional significance of this interaction would deserve further analysis.

4. Analysis of Pc as a Possible SUMO E3 Ligase for Sall Proteins in *Drosophila melanogaster*

In the previous section, we described the direct interaction between human SALL1 and CBX4 in mammalian cells, which could influence the SUMOylation and the levels of these proteins, as well as the formation of the Pc bodies. To further investigate these facts, we turned to *Drosophila* as a model system in an attempt to understand the developmental consequences of this interaction.

4.1. Salm and Pc partially colocalize *in vivo*

Previous experiments from the laboratory showed that Salm had a highly dynamic pattern of expression. During L3 Salm localized in the nucleus in a heterogeneous way (Fig. 47A-A'). Interestingly, when the larva was about to pupariate, Salm changed its distribution in the RG dramatically and decorated a ring around the nucleolus at the end of the L3 instar (Fig.

47B-B'). This localization is transient, and Salm localized again throughout the nucleus shortly after (Fig. 47C-C'). Through the work of different laboratory members, we discovered that this translocation was not a general behavior for other transcription factors such were Osa, HP1 or PEP (Jonatan Sanchez, PhD Thesis, 2009). We also discovered that the translocation coincided with the highest levels of ecdysone at the end of the L3 stages prior entering into pupariation (Soraya Curiel, unpublished results). However, the possible causes and consequences of the nucleolar translocation of Salm remain unknown. The rapidity of the translocation allowed us to speculate that it might be caused by a posttranslational modification of Salm.

In the course of our research, we observed a partial colocalization between Pc and Salm in the PG during L3 instar (Fig. 48). These two factors colocalize in certain foci in the polytene chromosomes of the PG of *Drosophila*, which could correspond to genes regulated by Pc. Interestingly, Pc weakly colocalized with Salm, at the end of L3, in the ring structure around the nucleolus (Fig. 48). These results are compatible with the interaction of Salm and Pc at this stage of development.

We also used cultured cells as a model to check the colocalization between Salm and Pc. As described before, we used *pUAST-CFP-HisMycSalm* construct in presence of *pAc-Gal4* to express Salm proteins in S2R+ *Drosophila* cells. Also in cultured cell, we showed partial colocalization between Salm and endogenous Pc (Fig. 48B-B'').

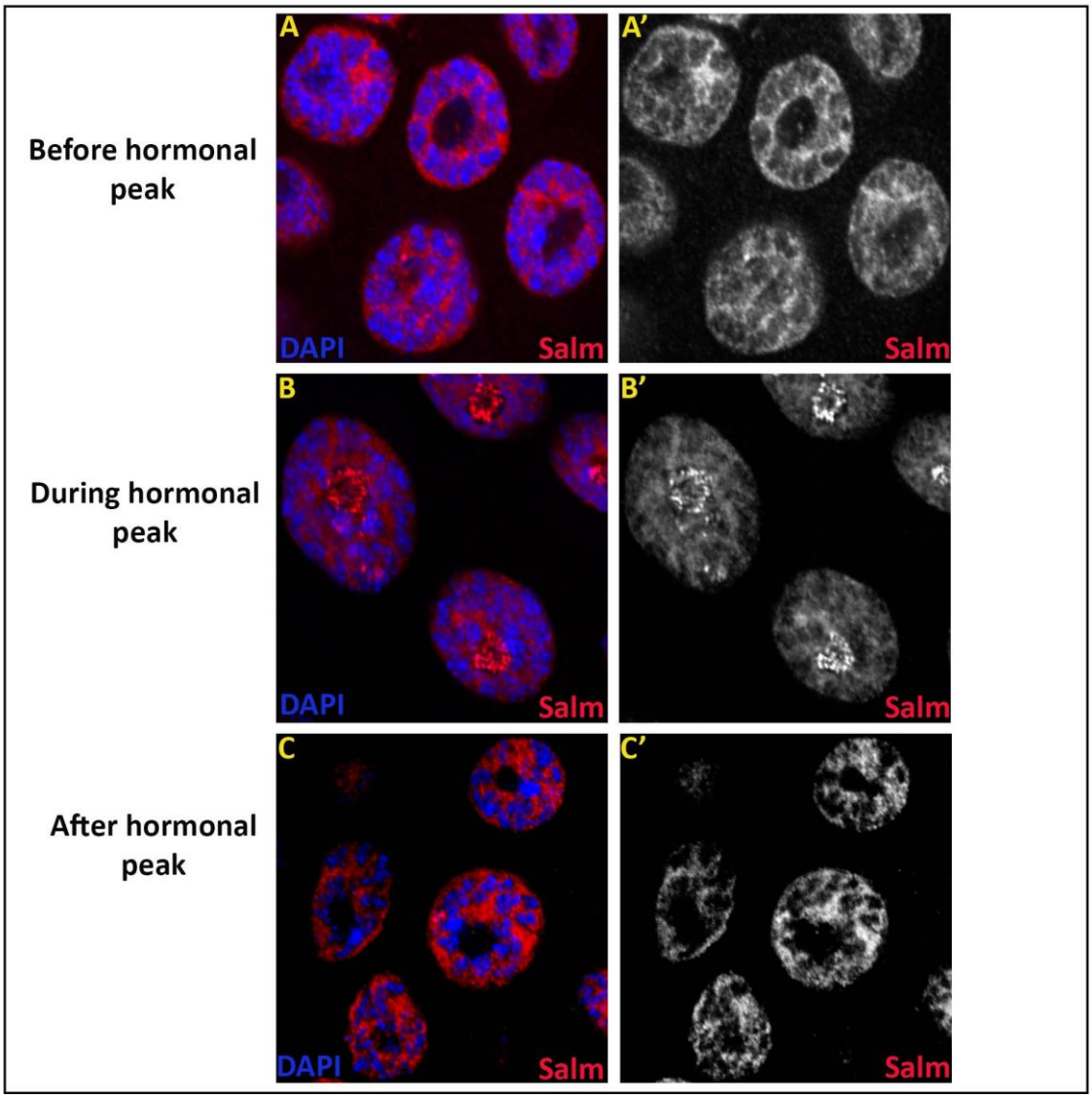


Figure 47. Localization of Salm in the PG during L3 instar. A-D'. Localization of Salm (red) in the PG cells before (**A**), during (**B**) and after (**C**) the hormonal peak at the end of the L3 instar. The translocation to the nucleolus (ring formation) is visible in **B-B**'. **A'-C'.** Red channel is presented in black and white independently. Nuclei were stained with DAPI (blue). Pictures were taken with a Leica DM IRE2 confocal microscope with a 63X objective.

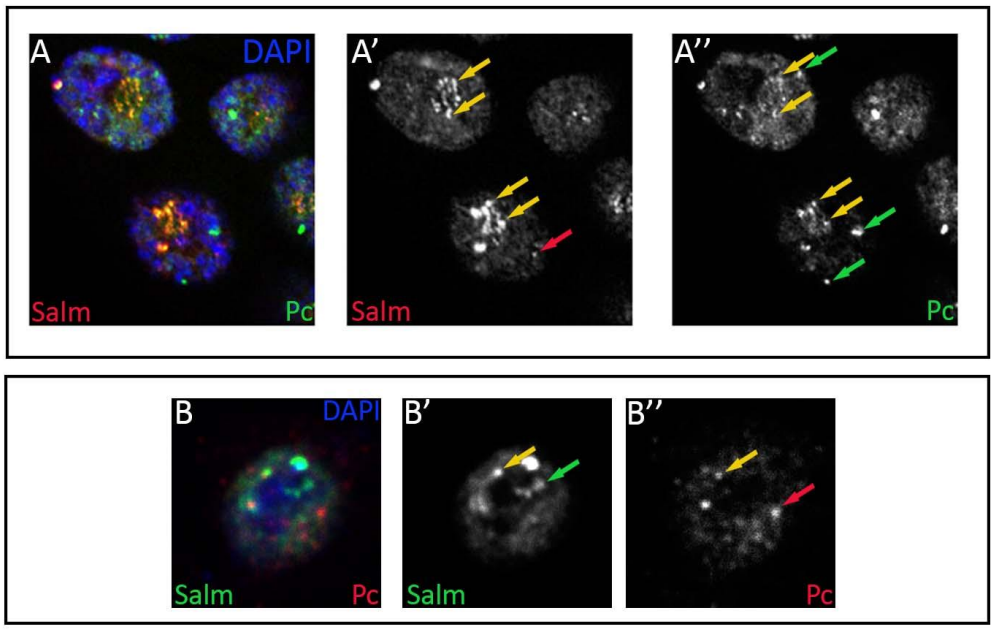


Figure 48. Partial colocalization between Salm and Pc. A-A''. Salm and Pc partially colocalize in PG cells during ring formation (indicated by yellow arrows) at the end of the L3 instar. **B-B''.** Salm and Pc partially colocalize in S2R+ cells transfected with CFP-Salm (yellow arrow). Green and red channels are shown independently in black and white. Nuclei were stained with DAPI (blue). Pictures were taken with a Leica DM IRE2 confocal microscope using a 63X objective.

4.2. Genetic interaction between *salm* and *Pc in vivo*

It was previously shown that SUMOylation has a role on Sall proteins during *Drosophila* wings development (Sánchez et al., 2010). We decided to investigate the nature of the Pc/Salm interaction, focusing on a possible role of Pc on Salm SUMOylation, using the adult wing as a model to perform the experiments (Barrio and De Celis, 2004). For this, we used the line Sal^{EPV}-*Gal4* as a driver to direct the overexpression or the silencing of different genes in the area where *sall* genes are normally expressed. This corresponds to the area spanning from the longitudinal vein LII to the region between veins LIV and LV comprising the intervein regions B to D (Fig. 49).

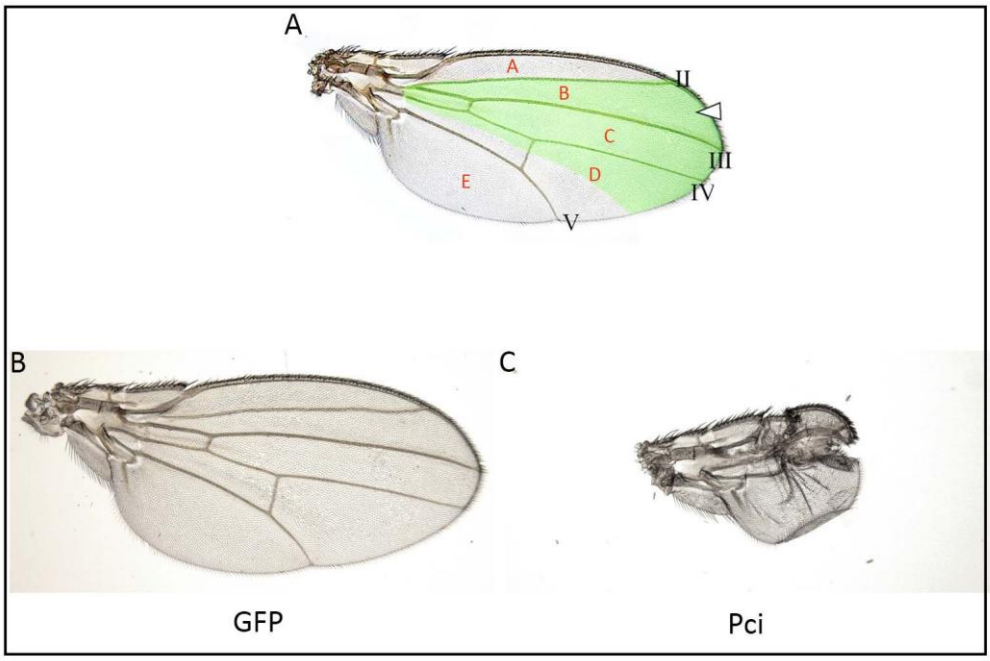


Figure 49. Silencing *Pc* in the wing. A. *Drosophila* wild type wing where the *Sal^{EPV}-Gal4* expression area is indicated (green). **B, C.** Pictures of adult wings of genotypes *Sal^{EPV}-Gal4>UAS-GFP* (GFP) and *Sal^{EPV}-Gal4>UAS-Pci* (Pci).

First we checked the phenotype resulting of knocking down *Pc* in the wing by crossing *Sal^{EPV}-Gal4* with *UAS-Pci*. As is show in Fig. 49, downregulation of *Pc* drastically reduced the wing areas B and C, resulting in the fusion of veins LII, LIII and LIV and in a significantly smaller wing area when compared to control wings.

It is known that Pc represses the expression of Ultrabithorax (Ubx) in the wing imaginal discs and that, in turn, Ubx transcriptionally represses the expression of *salm* and *salr* (de Navas et al., 2011; Weatherbee et al., 1998). Therefore, the de-repression of *Ubx* when silencing *Pc* precluded us to analyze the effect of Pc on Sall proteins at the post-transcriptional level. To avoid the transcriptional effect of Ubx on *sall* genes, we generated transgenic flies that silenced *Pc* and *Ubx* at the same time, *Sal^{EPV}-Gal4>UAS-Pci,UAS-Ubxi*. Silencing *Ubx* at the same time than *Pc* slightly but significantly rescued the size of the wings at 18^{o}C, where the effect of silencing *Pc* is milder than at 25^{o}C (Fig. 50).

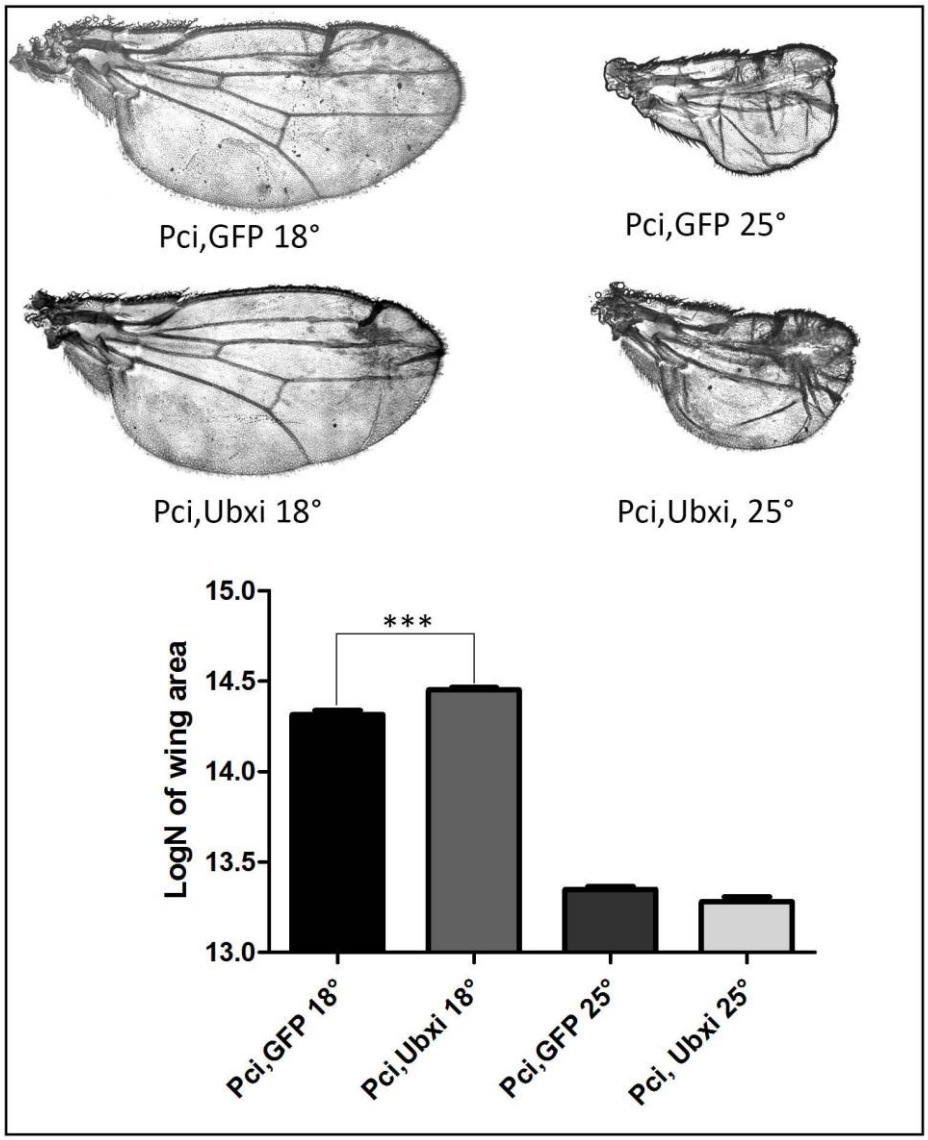


Figure 50. Silencing *Pc* in combination with *Ubx* in the wing. Pictures of adult wings of genotypes *Sal^{EPV}-Gal4>UAS-Pci,UAS-GFP* (Pci,GFP) and *Sal^{EPV}-Gal4>UAS-Pci,UAS-Ubxi* (Pci,Ubxi) at 18°C and 25°C. Below, the graph represents the logarithm of the average area of the wings of each genotype and temperature expressed in μm^2. Error bars indicate standard error of the mean. Three asterisks indicate $p<0.001$, calculated using Student's *t* test analysis.

The mild effect of silencing *Ubx* might reflect a poor efficiency of the *UAS-Ubxi* lines. In order to check the efficiency of silencing *Pc* and *Ubx* on *salm* expression, we immunostained wing imaginal discs with anti-Salm, anti-Pc and anti-Ubx antibodies (Fig. 51). These stainings were performed by Dr. C. Pérez in the laboratory. As expected, silencing *Pc* increased the levels of Ubx protein in the wing blade (Fig. 51B-B'''). *UAS-Ubxi* lines #37823 and #37825 slightly reduced Ubx intensity (Fig. 52). Surprisingly, Salm expression, although reduced, was still visible in the wing blade despite the presence of Ubx. This could maybe be due to the long perdurance of the protein.

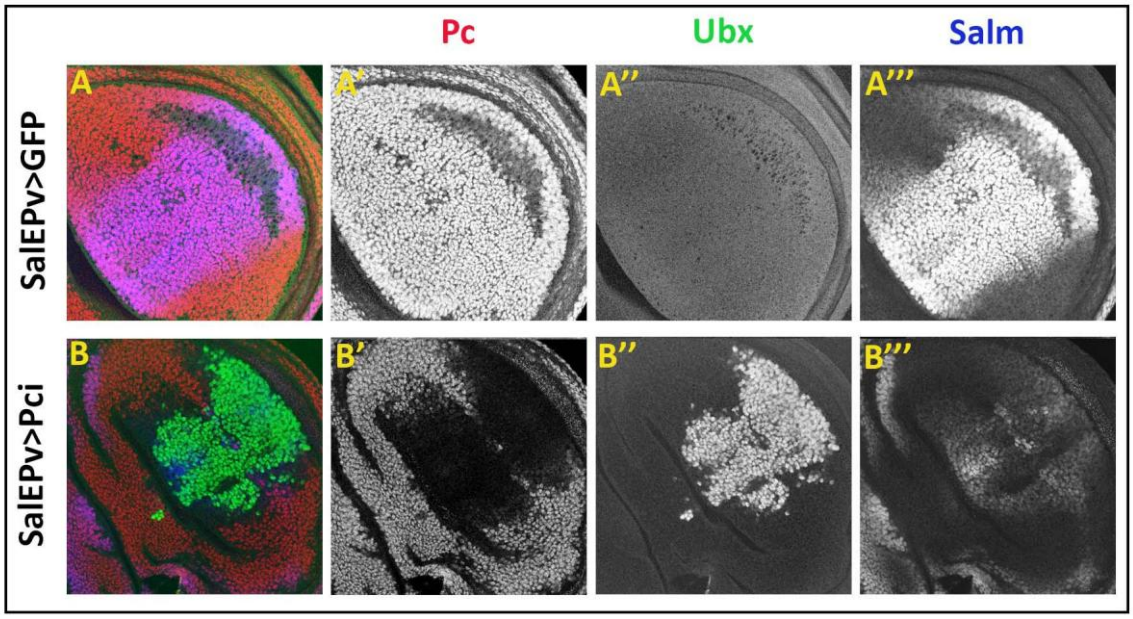


Figure 51. Effect of silencing *Pc* on Ubx and Salm expression. **A, B.** Wing discs of genotypes *Sal*^{EPV}*-Gal4>UAS-GFP* (SalEPv>GFP) and *Sal*^{EPV}*-Gal4>UAS-Pci* (SalEPv>Pci) showing Pc (red), Ubx (green) and Salm (blue) expression. **A'-B'''.** Red, green and blue channels are shown independently in black and white. Pictures were taken using a Leica DM IRE2 confocal microscope with a 63X objective.

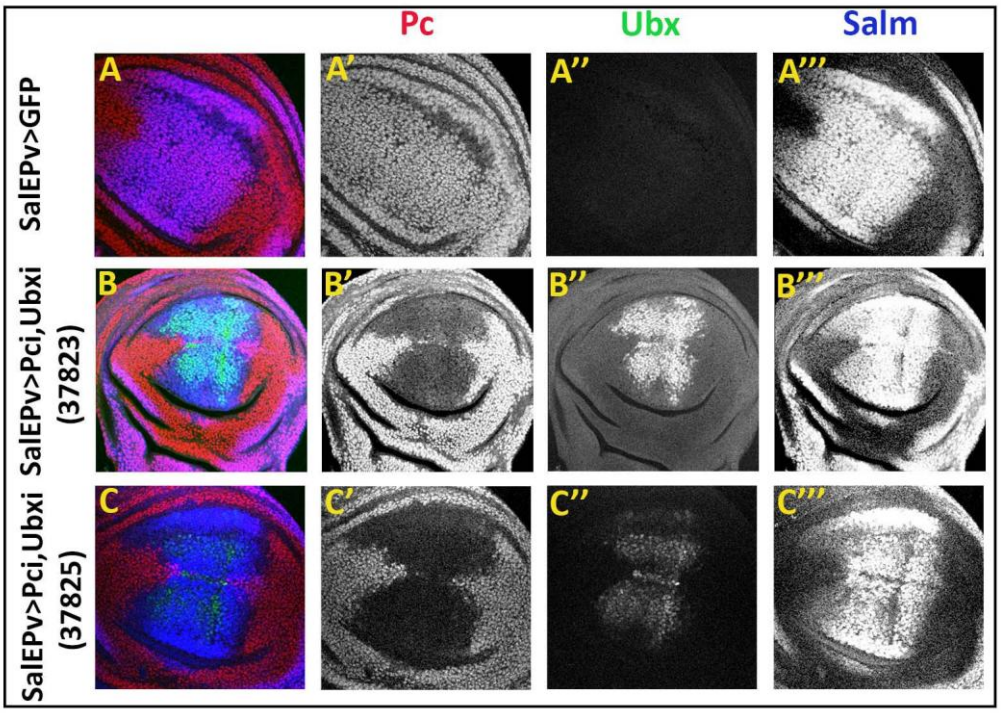


Figure 52. Comparison of different Ubxi lines. A-C. Wing discs of genotypes *Sal*^{EPV}*-Gal4>UAS-GFP* (**A-A'''**), *Sal*^{EPV}*-Gal4>UAS-Pci,UAS-Ubxi#37823* (**B-B'''**), *Sal*^{EPV}*-Gal4>UAS-Pci,UAS-Ubxi#37825* (**C-C'''**) showing Pc (red), Ubx (green) and Salm (blue) expression. **A'-C'''.** Red, green and blue channels are shown independently in black and white. Pictures were taken with a Leica DM IRE2 confocal microscope with a 63X objective.

4.2.1. Salm rescues the phenotype produced by *Pc* silencing

To check if Sall proteins could rescue the drastic phenotype generated by knocking *Pc* down, we generated the transgenic lines *Sal^{EPV}-Gal4>UAS-salm,UAS-Pci,UAS-Ubxi* and *Sal^{EPV}-Gal4>UAS-salr,UAS-Pci*. As shown in Fig. 53, overexpression of Salm rescued the phenotype of *Pci*. The phenotype of *Sal^{EPV}-Gal4>UAS-salm,UAS-Pci,UAS-Ubxi* wings was similar to the phenotype of the wings where *salm* was overexpressed in a WT background: we observed extravein formation in LII and LIII and a reduction in the wing area (Fig. 53), which was smaller than the control.

In the case of Salr, its overexpression did not rescue the phenotype produced by silencing *Pc* to the same extent than Salm, as *Sal^{EPV}-Gal4>UAS-salr,UAS-Pci* wings are similar to *Sal^{EPV}-Gal4>UAS-Pci* (Fig. 54). We cannot discard that the *UAS-salr* transgene was expressed at lower levels than *UAS-salm*, as these transgenes were generated by random insertion. Other *UAS-salr* lines would need to be analyzed in order to make a conclusion on the efficacy of Salm versus Salr to rescue the *Pci* phenotype.

4.2.2. Role of SUMOylation on *sall-Pc* genetic interaction

We hypothesized a possible implication of Pc in the Sall SUMOylation process as an E3 ligase. As we mentioned previously, ideally we should distinguish between the role of Pc on *sall* genes transcription and its putative role as E3 ligase. In order to do that, we used different transgenes. On one hand, we used the mutated forms of Salm and Salr in their respective SUMOylation sites, which makes them independent of endogenous SUMOylation; on the other hand, we used Smt3-Sall fusion proteins that mimic the constitutively SUMOylated forms of Salm and Salr. In the case of Salm, the Salm^{IKPD} mutant and the Smt3-Salm fusion contributed significantly less to the rescue of the *Pci* phenotype compared to the overexpression of the wild type protein (Fig. 53). Interestingly, there was not a significant difference between the overexpression of the Salm^{IKPD} mutant and the Smt3-Salm fusion, which indicated a complex relationship between Salm and SUMOylation in relation to its function in the wing.

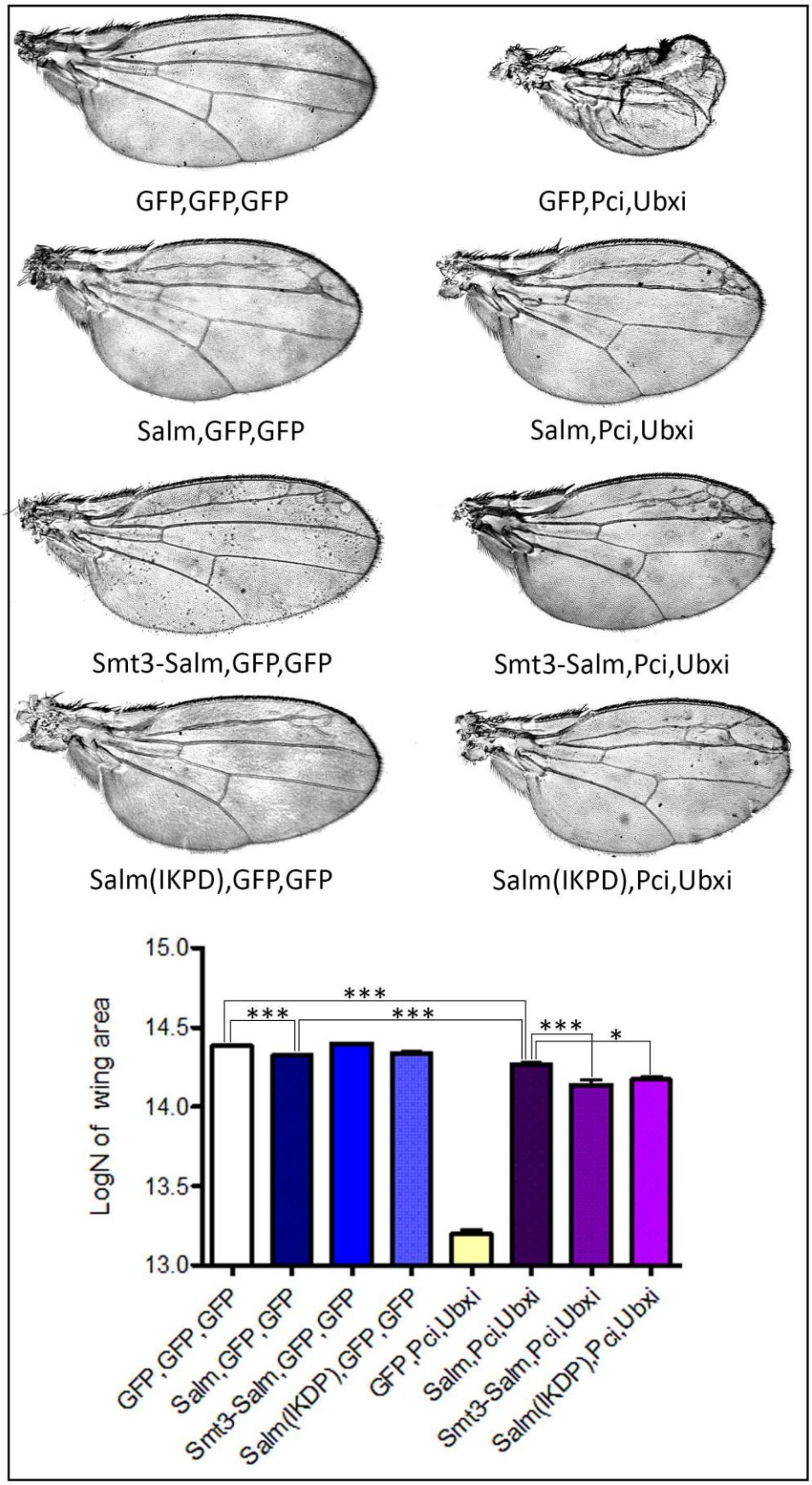


Figure 53. Analysis of the role of SUMOylation in *salm* and *Pc* genetic interaction. Pictures of adult wings of the indicated genotypes. Below, graphical representation of the logarithm of wing areas expressed in μm^2. Error bars indicated standard error of the mean. One asterisk indicates $p \leq 0.05$, three $p \leq 0.001$, calculated using Student's *t* test.

When we analyzed the set of experiments done with Salr we observed that the double mutant Salr^{IKEA,IKVA} and the fusion protein Smt3-Salr rescued the phenotype caused by silencing Pc, although to a lesser extent than the Salm counterparts. In this case, it is possible to establish a direct correlation among these different transgenes, as all of them were inserted in the same genomic region and presumably should have similar levels of expression. Interestingly, in this case the mutant form Salr^{IKEA,IKVA} showed a significant higher capacity to rescue *Pci* phenotype compared to Smt3-Salr (Fig. 54). These results were in agreement for a role of SUMOylation in inhibiting Salr function, as it was previously suggested (Sanchez et al., 2010). However, they did not support for a function for Pc on Sall SUMOylation during wing development.

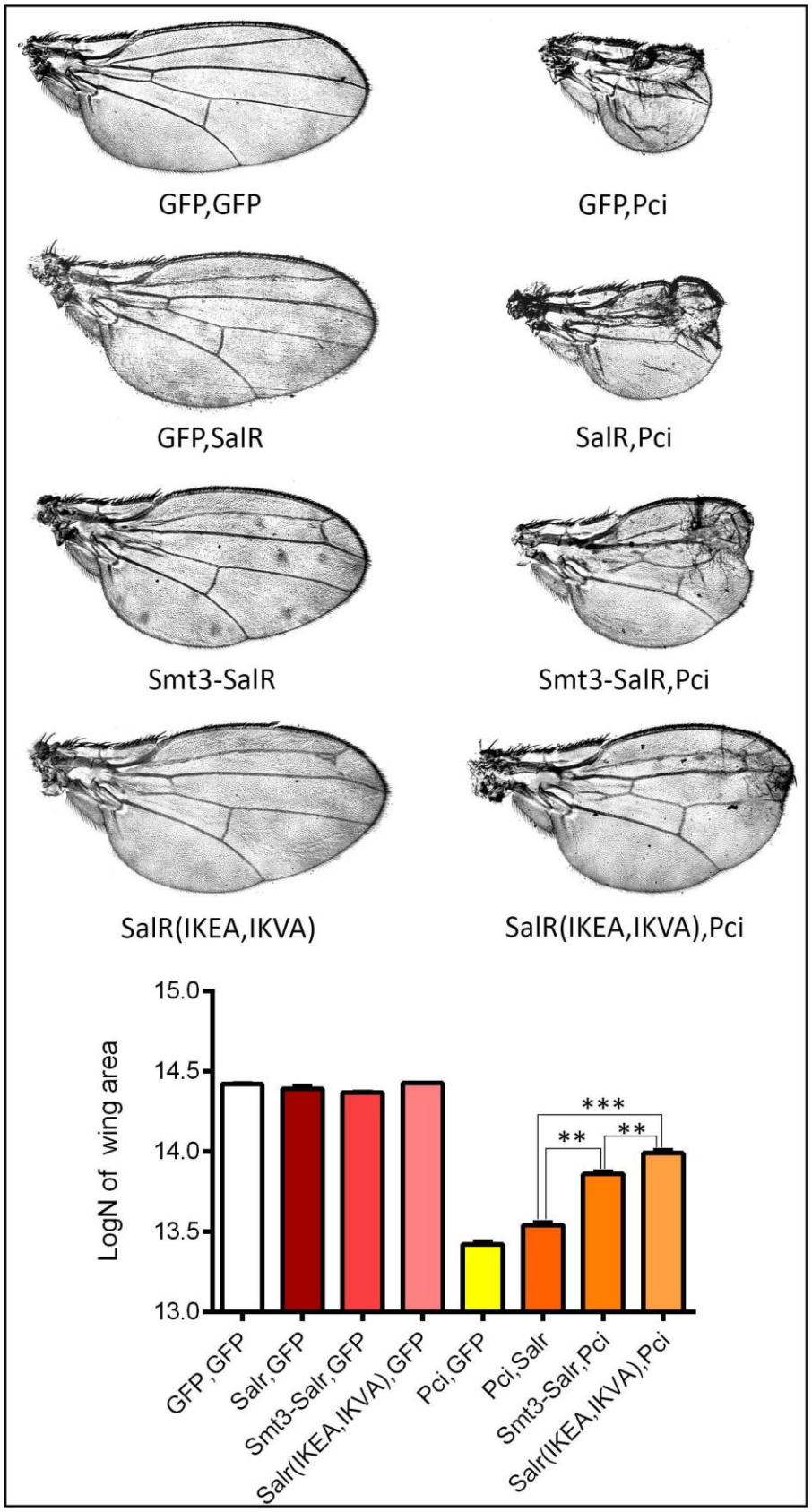
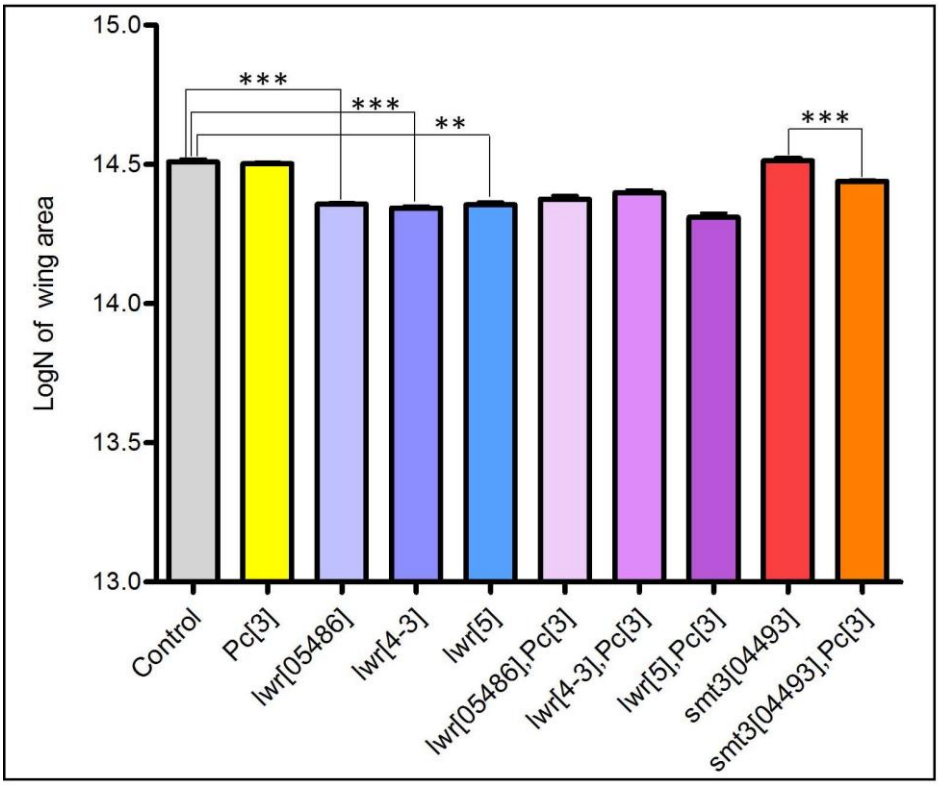


Figure 54. Analysis of the role of SUMOylation in *salr* and *Pc* genetic interaction. Pictures of adult wings of the indicated genotypes. Below, graphical representation of the logarithm of wing areas expressed in μm^2. Error bars indicated standard error of the mean. Two asterisks indicate $p \leq 0.01$, three asterisks indicate $p \leq 0.001$, calculated using Student's *t* test.

4.4.3. Interaction of *Pc* and the SUMOylation pathway: transheterozygous analysis

To further explore the relationship of Pc with the SUMOylation pathway during wing development, we performed a transheterozygous analysis. For this, we used the *Pc[3]* mutant combined in heterozygosity with the E2 enzyme mutants *lwr[05486]*, *lwr[4-3]* or *lwr[5]*. *Pc[3]* mutant wings did not varied in size with respect to the control (Fig. 55). However, *lwr[05486]*, *lwr[4-3]* and *lwr[5]* wings were significantly smaller than the control (Fig. 55). No major changes were found in size when combining *lwr* mutants with *Pc[3]*, these changes being in opposite directions with *lwr[4-3]* or *lwr[5]*. We also performed transheterozygosity experiments with *Pc[3]* mutant in combination with the *Smt3[04493]* mutant. *Smt3[04493]* wings size resulted very similar to the control, while the *Pc[3],Smt3[04493]* transheterozygous wings showed a significant reduction in relation to the *Pc[3]* or *Smt3[04493]* (Fig. 55). These results support an interaction between *Pc* and *smt3*, while the interaction with *lwr* is mild. However, we cannot discard a more subtle interaction that the trans-heterozygous analysis is not able to detect due to the relative abundance of the proteins that we are analyzing.



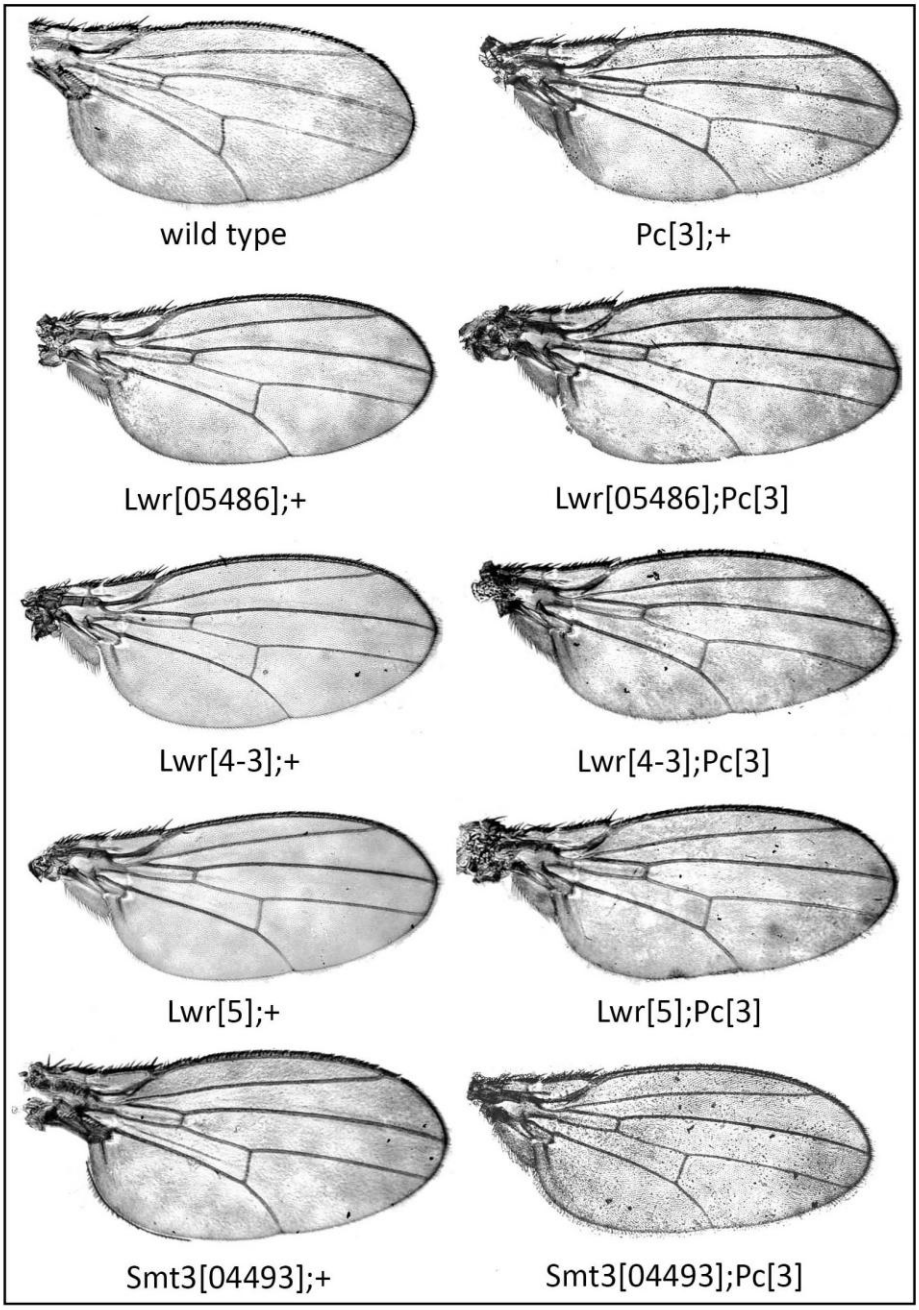


Figure 55. Analysis of a possible interaction between Pc and the SUMOylation machinery. Graphical representation of the logarithm of wing areas expressed in μm^2. Error bars indicate standard error of the mean. Two asterisk indicate $p \leq 0.01$, three asterisks indicate $p \leq 0.001$, calculated using Student's *t* test. Below, pictures of adult wings of the indicated genotypes.

We also used the trans-heterozygous analysis to further explore the relationship between *Pc* and *sall* genes in the wing. For that, we used the *Pc[3]* mutant combined in heterozygosity with *Df-5*, a deficiency that lacks both *salm* and *salr*. As mentioned above, wings from *Pc[3]* mutants did not differ significantly in size with respect to the control (Fig. 56). As reported previously, *Df-5* wings were smaller than control in a 76-78%, depending on the *Df-5* stock that was used (Fig. 56). However, in combination with *Pc[3]*, the wings significantly

recovered their size to a 85-94% of the control (Fig. 56). These results indicated a negative interaction of Pc on *sall* genes, which is in contrast with the possible de-repression of *Ubx* in the *Pc[3]* background.

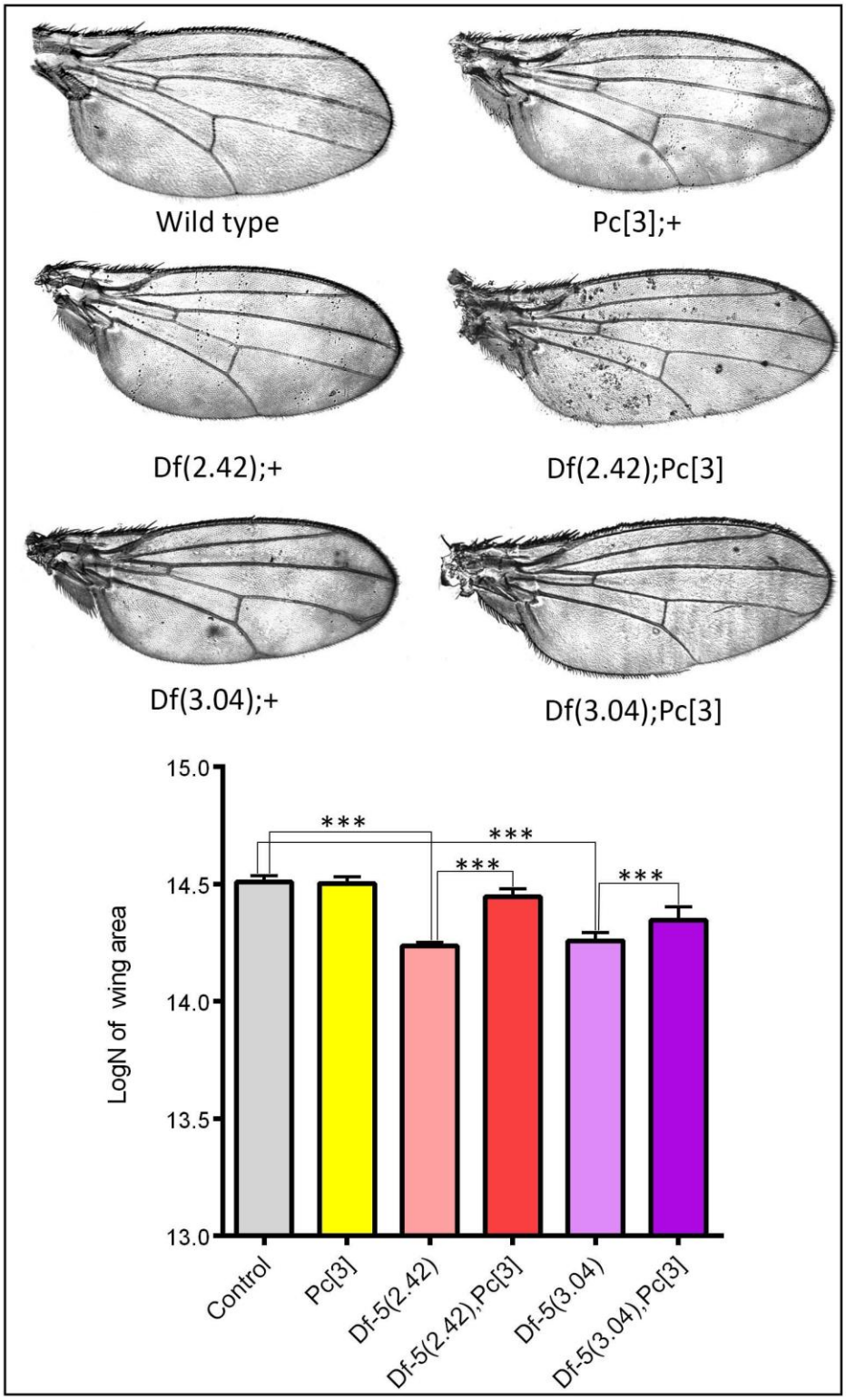


Figure 56. Analysis of a possible interaction between *Pc* and *sall* genes. Pictures of adult wings of the indicated genotypes. Below, graphical representation of the logarithm of wing areas expressed in μm^2. Error bars indicate standard error of the mean. Three asterisks indicate $p \leq 0.001$, calculated using Student's *t* test.

VI. DISCUSSION

Post-translational modifications (PTMs) control a wide range of physiological processes, contributing to extend the functionality and dynamics of eukaryotic proteomes. UbLs are involved in diverse biological and cellular processes. In contrast with Ubiquitination, SUMOylation is not necessarily involved in protein degradation. SUMO modulates the function of target proteins and is implicated in several biological processes such are DNA replication, DNA repair, chromatin remodeling, cell cycle, transcription, translation and stress response. In the last 20 years many different approaches were used to identify SUMOylated proteins that are important for the maintenance of the cellular equilibrium.

1. bioSmt3 methodology to analyze SUMOylation in *Drosophila*

Drosophila Smt3 is necessary throughout development. Differently from the situation in mammals and *S. cereviasie*, not many screenings have been performed in *Drosophila* to envision a complete scenario of the SUMOylated subproteome. Removing Smt3 from the PG blocks development at L3 larval instar and our long-term aim was to identify SUMOylated proteins that were involved in steroidogenesis. During this work we developed a new system, using a biotin tag at the N-terminal part of *Drosophila* Smt3, in order to identify SUMOylated protein in cell sand *in vivo*. The method was based on a technology developed in the laboratory of Dr. U. Mayor for the identification of ubiquitinated substrates *in vivo* (Franco et al., 2011). The first attempt, performed in cells and *in vivo,* using *pUAST-bioSmt3::BirA* did not allow us to get enough enrichment in the sample of biotinylated/SUMOylated proteins for a successful Mass Spectrometry identification. We think that this lack of SUMOylated material was mainly due to three reasons: (i) the relative abundance of SUMO-conjugated proteins is lower than the abundance of ubiquitinated ones; (ii) our vector contained a single *smt3* gene, while Franco and coworkers used a vector with seven repetitions of Ub; and (iii) the fusion protein bioSmt3::BirA was not completely processed in the cells, which impeded the isolation of enough material for Mass Spectrometry analysis after NeutrAvidin chromatography. Optimization of the technique by using the multicistronic vector *pAc5-STABLE2-Neo* (González et al., 2011) was successful. We performed a global screening in S2R+ *Drosophila* cells, after the demonstration of the efficacy of the system on a specific target, Ftz-f1, previously described to be SUMOylated *in vitro* (Talamillo et al., 2013).

The aim behind the development of the bioSmt3 system was to have an approach to perform a single-step pulldown under very stringent conditions, taking advantage of the strong affinity between biotin and NeutrAvidin. The efficacy of the system was demonstrated in cells through the validation of putative SUMOylated targets using exogenous proteins. More importantly, validation was also performed in endogenous conditions, revealing the sensitivity of the methodology. Using specific antibodies, we demonstrated the SUMOylation of Fax and Lam, which were not previously described to be conjugated to SUMO and those could be considered as targets for future investigations.

The *in vivo* analysis was performed using the anterior part of the larval body, which contains the imaginal discs, because it was reported the importance of SUMOylation in proliferating tissues during *Drosophila* development (Kanakousaki and Gibson, 2012). We observed a remarkable difference in terms of total number of putative targets when comparing the Mass Spectrometry list obtained from S2R+ cells (980) with the list obtained from larvae (92). In general, the recovery of material in cells is more direct and simple than the isolation of specific larval tissues. Also, in S2R+ cells, we expressed *lwr* at the same time than bioSmt3 to optimize the process of SUMO conjugation. These two aspect might be crucial to increase the yield of SUMOylated proteins.

To explore the functionality of bioSmt3, we analyzed two main aspects: (i) the localization of bioSmt3, that was mainly in nuclear bodies both in cells and *in vivo* as happens for the endogenous Smt3, and (ii) the capability to rescue the phenotype caused by removing *smt3* in the *Drosophila* PG. The *bioSmt3* transgene used in these experiments is insensitive to the dsRNA directed against the endogenous *smt3*, as we used a degenerated nucleotide sequence for the transgene (Sánchez et al., 2010). The complete rescue of the phenotype supported the validity of our methodology, indicating also that in *hs-Gal4>UAS-Smt3i,UAS-bioSmt3-T2A-BirA*^{opt} transgenic *Drosophila* the majority of Smt3 available is the bioSmt3 fusion. Interestingly, Tramtrack and Calmodulin appeared in the list of putative SUMOylated targets *in vivo*: these proteins are expressed in the PG and there is an indication of the SUMOylation of Tramtrack in the literature (Lehembre et al., 2000). The optimization of the recovery of material from tissues would be necessary for the identification the complete SUMOylated subproteome. Also, the collection of isolated PGs (or cerebral-PG complexes) will be advisable in order to identify the SUMOylated proteins that are implicated in steroidogenesis, reducing in this way the background from other tissues. However, we did not obtain results after dissecting 100 PGs per genotype when using the genetic combination *phm-Gal4>UAS-smt3i,UAS-bioSmt3::BirA.* This

results might improve by using the multicistronic vectors in presence of Lwr and by increasing the number of isolated PGs.

The GO terms obtained from the lists of proteins derived from the Mass Spectrometry experiments was what we expected from a SUMOylated subproteome, based on what it was described in other systems. We compared our lists among themselves (cultured cells *versus in vivo*) and with a list obtained from *Drosophila* embryos. The analysis of SUMOylated proteins in early development was performed using two different approaches: a two-step affinity purification strategy using His_6 and FLAG tags at the N-terminus of Smt3 and, under native condition, a single-step anti-FLAG immunopurification (Nie et al., 2009). The overlap between our lists and that of Nie et al. was significant, considering that the experiments were done using different biological systems. In all the lists, gene expression regulators were overrepresented, which was in agreement with the role of SUMOylation as a modulator of the transcriptional activity important during development.

2. New methodology for the analysis of bioUbL modifications in mammalian cells

The modularity of the *bioSmt3* vector allowed us to easily switch the promoter and to substitute the different modules. The screening performed in HEK 293FT cells using bioSUMO3 gave us the confirmation of the efficacy of the system. We identified and validated proteins conjugated to SUMO3 already described in mammalian cells. Based on that, we generated a collection of vectors including most of the UbL modifiers described until now. Importantly, this strategy could be very useful and easy to apply to any mammalian model system. First of all, almost any protein can be biotinylated *in vivo*; second, the procedure is simple, as only one-step pull down is required under stringent conditions avoiding more complicated biochemical procedures; third, the modified proteins can be easily visualized in cells or tissues using streptavidin fused to a fluorescent dye; fourth, the vectors can be used either in transient transfection or to generate stable cell lines; and fifth, the vectors give the possibility to express the corresponding E2 or E3 to enhance the modification of interest. Furthermore, the bioUbL methodology is very sensitive and allowed us to identify the modified proteins without using a previous heat shock treatment, which is usually done to stress cells and enhance protein

modifications. Despite the fact that the bioUbL methodology implies the overexpression of the modifiers, and this would be considered as less physiological than the identification of the endogenous counterpart using specific antibodies (Becker et al., 2013), the economical cost of the experiments is highly reduced. To overcome the caveat of the overexpression, a useful application of the bioUbL methodology would be the replacement of the endogenous *UbL* by the corresponding *bioUbL*, which would have endogenous levels of expression. Taken together, all these enumerated advantages place the bioUbL methodology among the first choices for researchers interested on UbL modification, based on its high versatility, reliability and accessibility to most laboratories.

3. SUMOylation of SALL proteins: Role of SUMOylation on SALL proteins localization

The link of human SALL proteins to the hereditary syndromes TBS and OS, justify the interest of studying this family of transcription factors, as well as the possible role that SUMOylation might have on these proteins function. In addition, numerous phenotypic studies in *Drosophila* Salm confirm the relevance of this family of proteins during development, as well as the role of SUMOylation on their function (Sánchez et al., 2010).

SALL proteins accumulate in nuclear foci that do not overlap with PML bodies, PcG bodies, or with speckles. In order to investigate on the nature of these bodies we checked the colocalization with SUMO proteins and we demonstrated that SALL1 and Salm overlap partially with bioSUMO1, bioSUMO2 or with the endogenous SUMO2/3 in mammalian cells. This partial colocalization differs from the one of the PML bodies, where SUMO is always present and could act as scaffold to build the nuclear bodies (reviewed by Nagai et al. 2011).

Previous results from the laboratory showed that SUMOylation influences Salm localization in HEK 293FT mammalian cells. By overexpressing exogenous Smt3, the size of Salm nuclear foci changed depending of the SUMOylation status of the protein, while the mutant form of Salm resulted insensitive to the presence of Smt3 (Sánchez et al., 2010). In this work we did not appreciate changes in the localization of human SALL1 or *Drosophila* Salm in presence of bioSUMOs. These results could be due to the fact that we overexpressed mammalian SUMOs instead of *Drosophila* Smt3: Smt3 is homolog to SUMO2/3 but the existing differences could

cause the homologs to act differently (Ureña et al 2015). Another hypothesis could be that, similar to what happen in the PcG bodies, the presence of SUMOs influences the size of the PcG bodies in opposite ways in human *versus Drosophila* cells (Gonzalez et al., 2014). The optimal condition to analyze Salm localization in presence of Smt3 would implicate the endogenous expression of Salm in *Drosophila* cells. However, there are not available cell lines that express endogenous Salm, while the overexpression of Salm is deleterious for the cells. For those reasons, we adopted U2OS cells as a model system. Our results showed that the localization of either SALL1 or Salm in these cells does not depend on their SUMOylation status, as mutants in acceptor lysines do not show apparent differences in localization with their WT counterpart. Therefore, our hypothesis is that the localization of SALL1 and Salm does not depend on their SUMOylation status, but on the presence of putative SIM motifs in their aminoacidic sequences. It would be important to demonstrate this hypothesis with future experiments and to determine whether certain factors need to be SUMOylated to be able to interact with SALL1/Salm.

4. E3 ligases involved in SALL SUMOylation

E3 ligases are enzymes that promote SUMOylation and can show specificity for their substrates. Since our long term aim is to understand the role of SUMOylation on SALL function, we focused on the identification of specific SUMO E3 ligases. Some lines of evidences directed our interest to the analysis of CBX4/Pc ligases. First, we identified CBX4 as an interactor of SALL1 through the BioID screen; second, we observed a partial colocalization of Salm and Pc in the PG cells. Biochemical evidences suggested that the interaction between SALL1 and CBX4 is direct, not requiring a third party. In addition, our results indicated that CBX4 could enhance SALL1 SUMOylation *in vitro.* The fact that SALL1 and CBX4 did not colocalize in the nuclear PcG bodies and the PLA confirmation of the interaction between these two transcription factors, supports the idea that these factors interact in the nucleoplasm. Interestingly, our experiments revealed an unexpected role of SALL1 on the regulation of CBX4: in presence of SALL1 we reported an increase in the SUMOylation of CBX4, as well as an increase in the number of PcG bodies. Assuming the relationship between these two proteins, our data could suggest and hypothetical E3 ligase activity for both proteins (Fig. 57). The increase in the number of PcG bodies in presence of SALL1 could also reveal a role for this transcription factor in the function of CBX4/Pc, as the PcG bodies are involved in the transcriptional repression of CBX4/Pc target

loci. The existence of modulators of the Pc function is an active area of research at the moment, and it would be very pertinent to confirm the role of SALL1 on this function. The role of SALL1 could be based on the stabilization of CBX4/Pc protein or on changes in the affinity of CBX4/Pc for its target DNA or interacting proteins. In any of these cases, it would be relevant to perform further investigation to elucidate the cross-relationship between SALL and CBX4/Pc proteins.

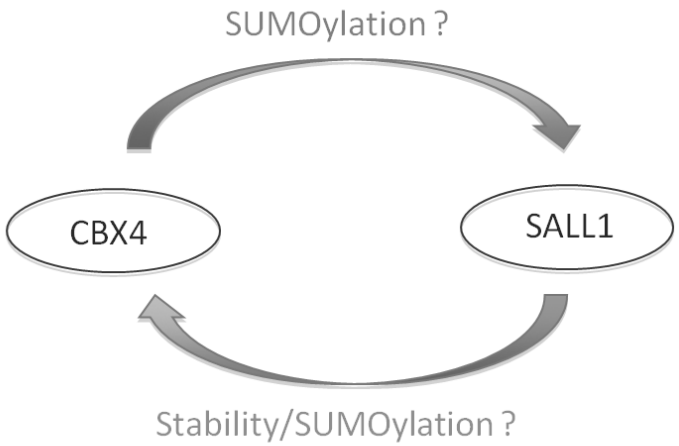


Figure 57. **Schematically representation of the hypothetical function of SALL1 and CBX4**. CBX4, as E3 ligase, could be involved in SALL1 SUMOylation. Our data suggest that, in turn, SALL1 could be involved in CBX4 SUMOylation and stabilization.

4.1. Analysis of the relationship between Drosophila Salm and Pc

Drosophila Salm and Pc, differently from their mammalian counterparts, show partial colocalization in *Drosophila* cultured cells and, more importantly, in steroidogenic tissues: Salm and Pc colocalize in the PG at the end of the L3 instar, when the level of ecdysone is high at the onset of pupariation. Then, Salm translocates to the nucleolus for a short lapse of time, being this a particular behavior not reported for other known transcription factor as Osa, HP1 or PEP. The direct interaction and the colocalization of SALL and Pc proteins would suggest a functional relationship between these factors. To explore this possibility, we turned to *Drosophila*, more specifically to the *Drosophila* wing as a model system. However, the transcriptional relationship between Salm and Pc hinder the analysis of the role of Pc on Salm at

the post-translational level: Pc represses the expression of *Ubx* in the wing, being this in turn a transcriptional repressor of *salm*. To overcome this difficulty, we performed several genetic experiments, including overexpression of Salm and Salr, either the WT or SUMO-modified and mutant forms, in a background where *Pc* and *Ubx* were silenced, as well as a transheterozygous analysis between mutants of *Pc*, the SUMOylation machinery and *sall*. From this battery of experiments, we could extract the following conclusions. First, the wings that overexpress WT *salm* in a *Pci* background are significantly smaller than wings that overexpress it in a WT background, indicating that the presence of Pc is relevant for the phenotype. Second, the overexpression of the constitutively SUMOylated or mutant forms of Salm rescues the *Pci* phenotype to a lesser extent than the WT counterpart, supporting the idea that the SUMOylation status if the protein is relevant for the rescue of the *Pci* phenotype. In this respect, it is puzzling that the constitutively SUMOylated and the mutant form of Salm behave in a similar way, despite the fact that they are modified in opposite ways in relation to SUMO. Fourth, the mutant form of Salr has a higher capacity to rescue the *Pci* phenotype than the constitutively SUMOylated and the WT forms. This result is in agreement to the previously reported enhancement of Salr activity when its SUMOylation is impeded (Sánchez et al., 2010). However, the mutation of the SUMOylation site should be irrelevant in a background where the putative E3 ligase is mutated. Fifth, our transheterozygous analysis shows a mild genetic interaction between *Pc* and the SUMOylation machinery during wing development. And sixth, *Pc* interacts genetically with *sall* genes in our heterozygous analysis in the wing. However, this interaction is counterintuitive, keeping in mind that the reduction of *Pc* significantly rescues the phenotype caused by the reduction of *sall* genes.

Taken together, these results indicate an interaction between Sall factors and Pc in the wing, although we cannot conclude whether this interaction is post-translational or not. We corroborated a previous observation in which SUMOylation influences the role of Sall proteins in the wing in opposite directions, underlying a major effect of SUMOylation on Salr *versus* Salm. Whether this difference depends on Pc is not clear and would deserve further studies.

VII. CONCLUSIONS

I. The bioSmt3 system, coupled with Mass Spectrometry is an efficient tool to identify SUMOylated proteins in cells and *in vivo* in *Drosophila*.

II. Use of multicistronic vector *pAc5-STABLE2-Neo* in bioSmt3 system is necessary for efficiency of methodology in cells and *in vivo.*

III. The sensitivity of the bioSmt3 system allows the detection of the SUMOylation of endogenous proteins.

IV. The bioUbL system is suitable for the study of UbL post-translational modifications in mammalian cells, as demonstrated with the bioSUMO3 and bioUFM1 studies.

V. SALL proteins are SUMOylated in mammalian cells. Human SALL1 SUMOylation resides in lysines K571, K592, K982 and/or K1086, while *Drosophila* Salm SUMOylation resides in lysine K641.

VI. SALL proteins localize in nuclear domains and partially colocalize with bioSUMO and endogenous SUMO.

VII. SALL1 interacts directly with SUMO2/3 in the nucleus.

VIII. CBX4 was identified as a putative interactor of human SALL1 in the BioID screening.

IX. SALL1 and CBX4 proteins interact directly, possibly in the nucleoplasm.

X. CBX4 could be implicated in SALL1 SUMOylation, while SALL1 could influence CBX4 SUMOylation and stability.

XI. The expression of SALL1 increases the number of PcG bodies in U2OS cells.

XII. *Drosophila* Salm and Pc colocalize *in vivo* in the PG cells, at the end of the third instar larva, during the hormonal peak.

XIII. *Drosophila salm* and *Pc* interact genetically during wing development. However, we could not demonstrate a post-translational role for Pc on Sall proteins in the wing.

VIII. APPENDICES

APPENDIX I. List of proteins conjugated by bioSmt3 detected in S2R+ *Drosophila* cells.

Protein names	Gene names	FBID KEY	ANNOTATION SYMBOL	Number of Proteins	Peptides	PEP	Average Intensity bioSmt3/Ctrl
	lwr	FBgn0010602	CG3018	1	16	0	4,20E+04
U1 small nuclear ribonucleoprotein 70 kDa	snRNP-U1-70K	FBgn0016978	CG8749	1	24	3,184E-100	2,43E+03
Otefin	Ote	FBgn0266420	CG5581	1	23	0	2,13E+03
	sip2	FBgn0031878	CG9188	6	37	0	2,13E+03
	CG4806	FBgn0260456	CG4806	3	38	0	2,10E+03
	Map60	FBgn0010342	CG1825	2	24	0	2,05E+03
RuvB-like helicase 1	pont	FBgn0040078	CG4003	1	20	0	1,61E+03
	CG9300	FBgn0036886	CG9300	1	25	0	1,53E+03
T-complex protein 1 subunit gamma	Cctgamma	FBgn0015019	CG8977	2	33	0	1,51E+03
	CG7839	FBgn0036124	CG7839	3	50	0	1,41E+03
	Sym	FBgn0037371	CG2097	1	42	0	1,39E+03
	CG11123	FBgn0033169	CG11123	1	27	0	1,19E+03
Eukaryotic translation initiation factor 2 subunit 3	Su(var)3-9	FBgn0263755	CG43664	5	23	3,1E-212	1,17E+03
Cytoplasmic dynein 1 intermediate chain	sw	FBgn0003654	CG18000	27	17	7,639E-195	1,15E+03
	CG4564	FBgn0263993	CG43736	2	23	0	1,14E+03
	Vap-33A	FBgn0029687	CG5014	4	9	8,284E-95	1,13E+03
	Incenp	FBgn0260991	CG12165	3	30	0	1,11E+03
	Tpr2	FBgn0032586	CG4599	3	22	0	1,09E+03
H/ACA ribonucleoprotein complex subunit 4	Nop60B	FBgn0259937	CG3333	5	19	2,3926E-92	1,06E+03
Eukaryotic translation initiation factor 3 subunit J	Adam	FBgn0027619	CG12131	1	15	2,646E-151	1,02E+03
	Uba2	FBgn0029113	CG7528	3	19	5,801E-141	9,77E+02
Lysine-specific demethylase NO66	CG2982	FBgn0266570	CG2982	2	25	4,173E-143	9,02E+02
Zinc finger protein CG2199	CG2199	FBgn0035213	CG2199	3	21	0	8,93E+02
	prod	FBgn0014269	CG18608	2	18	0	8,68E+02
DNA topoisomerase 2	Top2	FBgn0003732	CG10223	2	55	4,033E-303	8,56E+02

Serine/threonine-protein phosphatase 4 regulatory subunit 2	PPP4R2r	FBgn0030208	CG2890	7	28	0	8,08E+02
	kuk	FBgn0038476	CG5175	2	23	0	7,98E+02
Bifunctional glutamate/proline--tRNA ligase	Aats-glupro	FBgn0005674	CG5394	3	113	0	7,90E+02
Lamin Dm0	Lam	FBgn0002525	CG6944	1	61	0	7,86E+02
	qkr58E-1	FBgn0022986	CG3613	2	22	0	7,83E+02
	eIF4AIII	FBgn0037573	CG7483	2	17	1,387E-138	7,73E+02
H/ACA ribonucleoprotein complex subunit 2-like protein	NHP2	FBgn0029148	CG5258	1	8	2,4096E-26	7,66E+02
	smt3	FBgn0264922	CG4494	1	12	0	7,35E+02
40S ribosomal protein S3	RpS3	FBgn0002622	CG6779	3	17	9,954E-50	7,16E+02
Eukaryotic translation initiation factor 2 subunit 1	eIF-2alpha	FBgn0261609	CG9946	1	15	2,6329E-57	6,44E+02
	CG10496	FBgn0034631	CG10496	1	15	9,341E-156	6,26E+02
	Spp	FBgn0031260	CG11840	3	4	2,9596E-12	6,10E+02
Ran-binding protein 16	Ranbp16	FBgn0053180	CG33180	1	21	1,132E-103	5,96E+02
Transcription termination factor 2	lds	FBgn0002542	CG2684	2	34	6,548E-126	5,77E+02
Dynein heavy chain, cytoplasmic	Dhc64C	FBgn0261797	CG7507	3	81	0	5,73E+02
Kinesin-like protein Klp10A	Klp10A	FBgn0030268	CG1453	2	24	1,195E-302	5,68E+02
Nuclear cap-binding protein subunit 1	Cbp80	FBgn0022942	CG7035	2	18	2,2023E-92	5,45E+02
Proliferating cell nuclear antigen	PCNA	FBgn0005655	CG9193	2	9	1,2798E-68	5,43E+02
	CG10627	FBgn0036298	CG10627	1	16	6,722E-104	5,22E+02
	CG11376	FBgn0031216	CG11376	1	2	0,0010894	5,07E+02
	blanks	FBgn0035608	CG10630	2	17	3,23E-128	5,06E+02
	nop5	FBgn0026196	CG10206	4	19	2,009E-256	4,75E+02
Cytochrome b5	Cyt-b5	FBgn0264294	CG2140	4	2	1,3557E-39	4,73E+02
	ebd1	FBgn0035153	CG3371	1	24	7,946E-209	4,60E+02
	Saf-B	FBgn0039229	CG6995	4	31	1,281E-304	4,46E+02
	tum	FBgn0086356	CG13345	3	17	2,966E-254	4,43E+02
40S ribosomal protein SA	sta	FBgn0003517	CG14792	3	10	1,238E-202	4,42E+02
Ubiquitin specific protease 14	Usp14	FBgn0032216	CG5384	1	16	3,698E-130	4,34E+02
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	Ost48	FBgn0014868	CG9022	1	12	4,399E-160	4,28E+02
	pav	FBgn0011692	CG1258	4	22	6,075E-241	4,28E+02
	Hpr1	FBgn0037382	CG2031	1	25	8,151E-252	4,27E+02

	CG17494	FBgn0040011	CG17494	3	28	6,181E-193	4,21E+02
Nascent polypeptide-associated complex subunit alpha	Nacalpha	FBgn0086904	CG8759	1	7	3,873E-114	4,13E+02
RRP15-like protein	CG3817	FBgn0038275	CG3817	2	11	1,787E-111	4,02E+02
	CG10373	FBgn0032704	CG10373	4	5	0	3,87E+02
	Bruce	FBgn0266717	CG6303	4	41	2,035E-245	3,73E+02
	Hlc	FBgn0001565	CG1666	2	17	1,962E-85	3,68E+02
Replication factor C subunit 1	Gnf1	FBgn0004913	CG1119	1	23	8,553E-195	3,63E+02
	Su(var)2-10	FBgn0003612	CG8068	11	10	1,88E-114	3,55E+02
Eukaryotic peptide chain release factor subunit 1	eRF1	FBgn0036974	CG5605	2	9	1,114E-192	3,53E+02
	Rtnl1	FBgn0053113	CG33113	9	17	6,158E-246	3,53E+02
	Dlic	FBgn0030276	CG1938	2	16	4,484E-98	3,51E+02
60S ribosomal protein L4	RpL4	FBgn0003279	CG5502	2	14	7,4005E-69	3,49E+02
Protein suppressor of variegation 3-7	Su(var)3-7	FBgn0003598	CG8599	4	16	4,623E-271	3,48E+02
	CG4289	FBgn0037020	CG4289	1	9	3,941E-124	3,35E+02
	ADD1	FBgn0026573	CG8290	4	20	1,03E-128	3,33E+02
	nudC	FBgn0021768	CG9710	1	16	2,68E-189	3,30E+02
	CG7897	FBgn0266580	CG7897	3	42	3,872E-231	3,22E+02
Transcription factor kayak, isoforms D/sro	kay	FBgn0001297	CG33956	2	11	4,004E-218	3,21E+02
Protein tramtrack, beta isoform	ttk	FBgn0003870	CG1856	2	14	9,476E-267	3,18E+02
	DIP1	FBgn0024807	CG17686	15	4	1,077E-75	2,98E+02
	CG15816	FBgn0261570	CG42684	1	2	8,7724E-05	2,97E+02
60S acidic ribosomal protein P0	RpLP0	FBgn0000100	CG7490	1	11	1,136E-126	2,93E+02
	MAN1	FBgn0034962	CG3167	2	17	2,6934E-92	2,93E+02
	CG13773	FBgn0042092	CG13773	1	8	2,0311E-30	2,92E+02
	DMAP1	FBgn0034537	CG11132	1	19	5,375E-102	2,91E+02
	PpD3	FBgn0005777	CG8402	1	20	7,598E-121	2,91E+02
	Nup98-96	FBgn0039120	CG10198	3	28	3,242E-196	2,88E+02
Ubiquitin carboxyl-terminal hydrolase 7	Usp7	FBgn0030366	CG1490	1	30	4,247E-287	2,84E+02
SH3 domain-binding glutamic acid-rich protein homolog	Sh3beta	FBgn0035772	CG8582	1	12	1,515E-162	2,77E+02
Myeloid leukemia factor	Mlf	FBgn0034051	CG8295	9	7	3,378E-122	2,76E+02

Chromodomain-helicase-DNA-binding protein Mi-2 homolog	Mi-2	FBgn0262519	CG8103	16	38	9,767E-165	2,62E+02
	CG7946	FBgn0039743	CG7946	1	15	5,7526E-91	2,59E+02
Protein slender lobes	sle	FBgn0037810	CG12819	3	30	0	2,58E+02
	mei-38	FBgn0260986	CG14781	3	6	1,0991E-91	2,56E+02
Hsp90 co-chaperone Cdc37	Cdc37	FBgn0011573	CG12019	3	12	3,892E-185	2,56E+02
Myosin heavy chain, non-muscle	zip	FBgn0265434	CG15792	4	89	0	2,53E+02
	crp	FBgn0001994	CG7664	3	16	6,89E-132	2,51E+02
Probable elongation factor 1-beta	Ef1beta	FBgn0028737	CG6341	1	10	1,8374E-58	2,44E+02
E3 ubiquitin-protein ligase hyd	hyd	FBgn0002431	CG9484	2	45	0	2,44E+02
	E(var)3-9	FBgn0260243	CG11971	8	13	1,685E-196	2,43E+02
	CG1815	FBgn0039863	CG1815	4	24	6,914E-170	2,40E+02
	Dref	FBgn0015664	CG5838	3	14	3,7015E-81	2,38E+02
Chromosomal serine/threonine-protein kinase JIL-1	JIL-1	FBgn0020412	CG6297	3	17	7,452E-207	2,37E+02
Septin-1	Sep1	FBgn0011710	CG1403	2	8	1,806E-129	2,36E+02
T-complex protein 1 subunit alpha	T-cp1	FBgn0003676	CG5374	2	17	1,477E-119	2,32E+02
	CG17068	FBgn0031098	CG17068	1	13	5,365E-223	2,26E+02
26S protease regulatory subunit 4	Rpt2	FBgn0015282	CG5289	1	13	1,8506E-76	2,23E+02
	Mrp4	FBgn0263316	CG14709	11	23	3,3474E-98	2,22E+02
Dosage compensation regulator	mle	FBgn0002774	CG11680	24	30	4,912E-206	2,22E+02
Transcription factor AP-1	Jra	FBgn0001291	CG2275	1	7	1,5698E-37	2,21E+02
	Mtor	FBgn0013756	CG8274	4	69	0	2,20E+02
	smid	FBgn0016983	CG8571	5	17	2,6708E-98	2,18E+02
	stwl	FBgn0003459	CG3836	2	23	8,033E-212	2,15E+02
	Uba1	FBgn0023143	CG1782	1	31	0	2,15E+02
60S acidic ribosomal protein P2	RpLP2	FBgn0003274	CG4918	2	3	1,6568E-08	2,11E+02
	CG1024	FBgn0027514	CG1024	3	12	4,358E-98	2,11E+02
	CG3287	FBgn0265003	CG44154	2	14	3,164E-239	2,07E+02
	Rpn1	FBgn0028695	CG7762	1	19	3,4961E-54	2,04E+02
	CG9123	FBgn0030629	CG9123	25	11	2,2589E-52	2,03E+02
Probable arginine--tRNA ligase, cytoplasmic	Aats-arg	FBgn0027093	CG9020	1	21	3,2894E-68	1,98E+02
	Su(var)2-HP2	FBgn0026427	CG12864	3	26	3,66E-140	1,96E+02

Zinc finger protein hangover	hang	FBgn0026575	CG32575	4	19	9,062E-178	1,95E+02
	coil	FBgn0033265	CG8710	4	18	9,1873E-79	1,94E+02
Actin-42A	Act42A	FBgn0000043	CG12051	1	18	0	1,93E+02
	CG11092	FBgn0027537	CG11092	4	21	2,301E-123	1,92E+02
	CG12304	FBgn0036515	CG12304	4	9	2,918E-24	1,92E+02
Zinc finger protein ush	ush	FBgn0003963	CG2762	2	12	2,2434E-90	1,88E+02
Kinesin heavy chain	Khc	FBgn0001308	CG7765	2	28	1,692E-178	1,87E+02
	CG13730	FBgn0052176	CG32176	2	9	1,1504E-91	1,84E+02
	Pcf11	FBgn0264962	CG10228	3	25	1,035E-132	1,84E+02
Phosphatidylinositol-binding clathrin assembly protein LAP	lap	FBgn0086372	CG2520	4	10	3,051E-100	1,82E+02
	CG7261	FBgn0027509	CG7261	2	18	5,9973E-72	1,82E+02
	CG1316	FBgn0035526	CG1316	1	7	3,4675E-32	1,82E+02
	mrt	FBgn0039507	CG3361	2	17	2,502E-137	1,81E+02
F-box-like/WD repeat-containing protein ebi	ebi	FBgn0263933	CG4063	3	10	1,463E-256	1,80E+02
Bifunctional protein BirA	birA			1	27	0	1,79E+02
26S proteasome non-ATPase regulatory subunit 1	Rpn2	FBgn0028692	CG11888	2	20	7,528E-114	1,76E+02
T-complex protein 1 subunit delta	CG5525	FBgn0032444	CG5525	1	10	5,299E-65	1,74E+02
Cleavage and polyadenylation specificity factor subunit 1	Cpsf160	FBgn0024698	CG10110	2	18	9,2798E-64	1,74E+02
	CG7262	FBgn0038274	CG7262	1	19	1,628E-123	1,71E+02
Centrosome-associated zinc finger protein CP190	Cp190	FBgn0000283	CG6384	1	19	1,25E-252	1,71E+02
Casein kinase I isoform alpha	CkIalpha	FBgn0015024	CG2028	3	9	2,3894E-22	1,71E+02
	CG1416	FBgn0032961	CG1416	2	9	7,5123E-97	1,70E+02
Polycomb protein PHO	pho	FBgn0002521	CG17743	1	6	1,5549E-16	1,69E+02
Putative fatty acyl-CoA reductase CG8306	CG8306	FBgn0034142	CG8306	2	10	1,1121E-66	1,68E+02
Surfeit locus protein 6 homolog	Surf6	FBgn0038746	CG4510	1	8	2,8026E-15	1,67E+02
Enhancer of mRNA-decapping protein 4 homolog	Ge-1	FBgn0032340	CG6181	1	27	6,6692E-78	1,65E+02
	CG2064	FBgn0033205	CG2064	5	11	1,8442E-37	1,63E+02
	CG7265	FBgn0038272	CG7265	1	9	2,888E-185	1,62E+02
	cg	FBgn0000289	CG8367	13	12	4,1E-48	1,61E+02
Serine-arginine protein 55	B52	FBgn0004587	CG10851	9	7	1,8171E-37	1,61E+02

Longitudinals lacking protein, isoforms F/I/K/T	lola	FBgn0005630	CG12052	1	9	1,104E-194	1,58E+02
	tho2	FBgn0031390	CG31671	3	27	1,049E-264	1,57E+02
	CG17090	FBgn0035142	CG17090	4	15	8,6649E-76	1,56E+02
	Apc	FBgn0015589	CG1451	5	5	3,893E-14	1,55E+02
Heterogeneous nuclear ribonucleoprotein 27C	Hrb27C	FBgn0004838	CG10377	10	5	1,074E-128	1,54E+02
	unc-45	FBgn0010812	CG2708	2	18	8,1125E-89	1,53E+02
	zf30C	FBgn0270924	CG3998	2	12	7,6031E-44	1,52E+02
	wol	FBgn0261020	CG7870	1	6	5,3243E-20	1,51E+02
	BtbVII	FBgn0263108	CG43365	3	5	1,3566E-14	1,51E+02
Bipolar kinesin KRP-130	Klp61F	FBgn0004378	CG9191	4	31	1,26E-114	1,50E+02
	CG11107	FBgn0033160	CG11107	1	13	9,9659E-83	1,47E+02
	CG7033	FBgn0030086	CG7033	2	31	0	1,45E+02
	Rs1	FBgn0021995	CG2173	2	15	7,5499E-86	1,44E+02
Cofilin/actin-depolymerizing factor homolog	tsr	FBgn0011726	CG4254	1	11	2,048E-169	1,43E+02
Elongation factor 1-gamma	Ef1gamma	FBgn0029176	CG11901	1	13	7,22E-204	1,42E+02
	msps	FBgn0027948	CG5000	5	38	1,789E-103	1,41E+02
	Nup358	FBgn0039302	CG11856	3	39	5,295E-205	1,41E+02
F-actin-capping protein subunit beta	cpb	FBgn0011570	CG17158	1	8	1,1E-53	1,38E+02
Protein extra bases	kra	FBgn0250753	CG2922	1	10	4,8781E-51	1,37E+02
	east	FBgn0261954	CG4399	4	19	0	1,35E+02
	p53	FBgn0039044	CG33336	5	7	1,0992E-21	1,35E+02
Protein held out wings	how	FBgn0264491	CG10293	6	8	2,2933E-27	1,34E+02
Hsc70-interacting protein 2	HIP	FBgn0260484	CG32789	4	9	6,4516E-66	1,33E+02
	mod(mdg4)	FBgn0002781	CG32491	1	10	1,737E-142	1,32E+02
Clathrin heavy chain	Chc	FBgn0000319	CG9012	1	24	9,279E-102	1,30E+02
	Df31	FBgn0022893	CG2207	1	15	3,013E-262	1,28E+02
	CG8569	FBgn0033752	CG8569	1	10	8,7411E-27	1,27E+02
	VhaM8.9	FBgn0037671	CG8444	1	7	1,304E-24	1,27E+02
	bon	FBgn0023097	CG5206	3	12	1,7611E-86	1,27E+02
THO complex subunit 5	thoc5	FBgn0034939	CG2980	1	7	2,0078E-68	1,26E+02

Lysine-specific demethylase lid	lid	FBgn0031759	CG9088	1	20	2,618E-106	1,25E+02
	CG12567	FBgn0039958	CG12567	8	6	1,1117E-25	1,25E+02
Ribonucleoside-diphosphate reductase large subunit	RnrL	FBgn0011703	CG5371	1	16	3,105E-119	1,24E+02
GATA-binding factor A	pnr	FBgn0003117	CG3978	4	5	1,6737E-73	1,23E+02
	Aats-asp	FBgn0002069	CG3821	2	14	0	1,21E+02
DNA replication licensing factor Mcm5	Mcm5	FBgn0017577	CG4082	1	14	4,0481E-42	1,19E+02
	Cpsf73	FBgn0261065	CG7698	1	13	5,1989E-63	1,19E+02
Protein odr-4 homolog	CG10616	FBgn0036286	CG10616	3	6	3,5303E-96	1,19E+02
	MESR4	FBgn0034240	CG4903	4	14	5,3546E-80	1,18E+02
Polynucleotide 5-hydroxyl-kinase NOL9	CG8414	FBgn0034073	CG8414	1	15	2,015E-171	1,18E+02
	Aats-thr	FBgn0027081	CG5353	4	18	1,6456E-70	1,17E+02
	homer	FBgn0025777	CG11324	5	11	1,4405E-32	1,17E+02
	CG8331	FBgn0033906	CG8331	5	2	4,1636E-06	1,17E+02
Protein painting of fourth	Pof	FBgn0035047	CG3691	1	8	7,7358E-30	1,16E+02
Eukaryotic translation initiation factor 4E	eIF-4E	FBgn0015218	CG4035	4	6	5,2967E-27	1,16E+02
	CG13096	FBgn0032050	CG13096	2	12	7,0451E-50	1,15E+02
	mbm	FBgn0086912	CG11604	1	5	2,6948E-26	1,13E+02
Nuclear pore complex protein Nup214	Nup214	FBgn0010660	CG3820	1	19	1,452E-102	1,12E+02
	CG5482	FBgn0034368	CG5482	1	9	1,5915E-27	1,12E+02
	l(2)35Df	FBgn0001986	CG4152	1	19	3,0464E-53	1,12E+02
	Spindly	FBgn0031549	CG15415	2	14	4,5584E-49	1,11E+02
	Rpn6	FBgn0028689	CG10149	2	10	4,6638E-44	1,08E+02
	CG4769	FBgn0035600	CG4769	1	3	2,2799E-08	1,08E+02
	tou	FBgn0033636	CG10897	4	25	1,75E-100	1,07E+02
DNA ligase 1	DNA-ligI	FBgn0262619	CG5602	1	11	2,7478E-71	1,06E+02
	CG12288	FBgn0032620	CG12288	2	6	7,2799E-94	1,06E+02
Nuclear export mediator factor NEMF homolog	Clbn	FBgn0259152	CG11847	3	14	3,0599E-43	1,06E+02
	Lhr	FBgn0034217	CG18468	3	9	7,5046E-43	1,05E+02
	Sin	FBgn0028402	CG10582	5	12	2,7274E-47	1,03E+02
	Sin3A	FBgn0022764	CG8815	12	19	2,2528E-90	1,03E+02

	Rpt1	FBgn0028687	CG1341	2	14	1,9657E-69	1,02E+02
	Elf	FBgn0020443	CG6382	2	13	1,7252E-57	1,02E+02
	Sec61alpha	FBgn0086357	CG9539	2	3	3,6456E-07	1,02E+02
DNA topoisomerase 1	Top1	FBgn0004924	CG6146	5	12	1,0715E-62	1,01E+02
	MSBP	FBgn0030703	CG9066	1	9	9,2967E-64	1,01E+02
FK506-binding protein 59	FKBP59	FBgn0029174	CG4535	1	7	3,2859E-19	1,00E+02
60S ribosomal protein L22	RpL22	FBgn0015288	CG7434	1	11	3,149E-102	9,98E+01
	D19A	FBgn0022935	CG10269	4	13	1,524E-113	9,92E+01
	Nup133	FBgn0039004	CG6958	11	14	1,1733E-74	9,77E+01
	cher	FBgn0014141	CG3937	11	47	4,499E-161	9,74E+01
	CG7246	FBgn0030081	CG7246	6	12	4,0355E-35	9,64E+01
60S ribosomal protein L14	RpL14	FBgn0017579	CG6253	3	4	1,2078E-12	9,63E+01
ATP-dependent RNA helicase WM6	Hel25E	FBgn0014189	CG7269	2	10	6,1568E-29	9,58E+01
	Nup107	FBgn0027868	CG6743	12	14	2,0001E-56	9,49E+01
	CG8436	FBgn0037670	CG8436	1	7	1,9162E-59	9,45E+01
Adenylyl cyclase-associated protein	capt	FBgn0261458	CG33979	3	6	1,3898E-69	9,40E+01
	Gdi	FBgn0004868	CG4422	2	12	6,8992E-40	9,40E+01
	CG12301	FBgn0036514	CG12301	9	10	7,8819E-64	9,33E+01
Integrator complex subunit 3 homolog	IntS3	FBgn0262117	CG17665	2	10	2,2802E-25	9,32E+01
Sterile alpha and TIR motif-containing protein 1	Ect4	FBgn0262579	CG43119	6	16	9,352E-188	9,24E+01
	mdy	FBgn0004797	CG31991	1	18	3,3899E-90	9,18E+01
	ste24a	FBgn0034176	CG9000	1	6	1,7359E-41	9,12E+01
Protein FAM50 homolog	CG12259	FBgn0039557	CG12259	1	5	6,4654E-23	9,10E+01
	row	FBgn0033998	CG8092	3	15	4,8755E-59	9,03E+01
Eukaryotic translation initiation factor 3 subunit M	Tango7	FBgn0033902	CG8309	1	6	2,1204E-35	8,96E+01
Chromatin-remodeling complex ATPase chain Iswi	Iswi	FBgn0011604	CG8625	2	14	6,5658E-42	8,91E+01
Transcription factor GAGA	Trl	FBgn0013263	CG33261	14	4	1,6193E-37	8,88E+01
	CG8258	FBgn0033342	CG8258	1	24	1,358E-145	8,82E+01
Protein suppressor of hairy wing	su(Hw)	FBgn0003567	CG8573	1	9	2,1586E-51	8,81E+01
	crol	FBgn0020309	CG14938	5	12	6,0847E-42	8,79E+01

	Dp1	FBgn0027835	CG5170	4	35	6,532E-140	8,71E+01
	CG5787	FBgn0032454	CG5787	1	11	6,7342E-44	8,69E+01
	CG8858	FBgn0033698	CG8858	1	22	2,6756E-93	8,61E+01
Eukaryotic translation initiation factor 3 subunit	eIF3-S10	FBgn0037249	CG9805	2	29	5,138E-100	8,45E+01
	CG8552	FBgn0031990	CG8552	2	12	1,748E-174	8,40E+01
	Nup75	FBgn0034310	CG5733	8	12	1,1914E-32	8,34E+01
	Past1	FBgn0016693	CG6148	5	14	3,6651E-33	8,31E+01
	ade5	FBgn0020513	CG3989	4	6	4,0633E-16	8,16E+01
Annexin;Annexin-B10	AnxB10	FBgn0000084	CG9579	7	12	2,2905E-42	8,14E+01
	pod1	FBgn0029903	CG4532	4	11	4,5864E-67	8,00E+01
Protein dopey-1 homolog	CG15099	FBgn0034400	CG15099	1	24	6,41E-114	7,97E+01
	Aats-val	FBgn0027079	CG4062	2	13	2,2689E-50	7,82E+01
	dgt1	FBgn0039710	CG18041	1	6	6,518E-56	7,74E+01
Serendipity locus protein delta	Sry-delta	FBgn0003512	CG17958	1	5	2,3912E-14	7,68E+01
GTP-binding protein CG1354	CG1354	FBgn0030151	CG1354	4	8	1,9782E-26	7,66E+01
	Rpn5	FBgn0028690	CG1100	1	15	1,8891E-43	7,62E+01
Eukaryotic translation initiation factor 3 subunit C	eIF3-S8	FBgn0034258	CG4954	1	17	4,9692E-52	7,61E+01
Ribonucleoprotein RB97D	Rb97D	FBgn0004903	CG6354	5	7	3,4768E-17	7,39E+01
	CG10462	FBgn0032815	CG10462	4	11	3,4916E-88	7,26E+01
	CG12608	FBgn0030630	CG12608	10	9	2,5773E-67	7,21E+01
Probable glutamine--tRNA ligase	Aats-gln	FBgn0027090	CG10506	1	12	3,1384E-56	7,14E+01
	alt	FBgn0038535	CG18212	3	17	2,392E-182	7,10E+01
	ATPCL	FBgn0020236	CG8322	4	13	1,9241E-80	7,09E+01
Alanine--tRNA ligase, cytoplasmic	Aats-ala	FBgn0027094	CG13391	5	12	4,9086E-88	7,09E+01
	CG9740	FBgn0037669	CG9740	1	4	1,1813E-11	7,09E+01
	Chro	FBgn0044324	CG10712	3	10	1,0915E-90	7,07E+01
	CG3071	FBgn0023527	CG3071	2	7	1,842E-15	7,05E+01
	CG1832	FBgn0032979	CG1832	3	3	6,0169E-35	7,05E+01
	CG3815	FBgn0029861	CG3815	1	9	1,002E-62	7,04E+01
	CG2662	FBgn0024993	CG2662	2	6	7,9515E-28	7,03E+01

UPF0505 protein CG8202	CG8202	FBgn0037622	CG8202	2	10	1,5519E-27	6,94E+01
	CG6686	FBgn0032388	CG6686	4	10	2,9787E-37	6,91E+01
La protein homolog	La	FBgn0011638	CG10922	3	10	3,8733E-34	6,86E+01
	CG12325	FBgn0033557	CG12325	1	7	2,3527E-22	6,81E+01
	CG3163	FBgn0034961	CG3163	1	8	8,2945E-22	6,77E+01
FACT complex subunit Ssrp1	Ssrp	FBgn0010278	CG4817	22	13	6,7142E-44	6,75E+01
Tropomyosin-1, isoforms 9A/A/B	Tm1	FBgn0003721	CG4898	15	7	5,3456E-39	6,73E+01
	Aats-tyr	FBgn0027080	CG4561	1	10	4,3312E-29	6,69E+01
	CG6241	FBgn0037792	CG6241	1	9	1,2879E-50	6,68E+01
Catalase	Cat	FBgn0000261	CG6871	1	9	2,4232E-81	6,65E+01
Eukaryotic initiation factor 4A	eIF-4a	FBgn0001942	CG9075	4	7	1,8081E-86	6,63E+01
	REG	FBgn0029133	CG1591	1	6	7,6905E-15	6,62E+01
	CG1703	FBgn0030321	CG1703	22	18	1,3279E-51	6,59E+01
NAD-dependent protein deacetylase Sirt2	Sirt2	FBgn0038788	CG5085	2	5	6,975E-284	6,56E+01
Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type	Ca-P60A	FBgn0263006	CG3725	3	16	4,4201E-81	6,54E+01
Annexin-B9;Annexin	AnxB9	FBgn0000083	CG5730	5	7	2,3651E-18	6,53E+01
Probable nucleolar GTP-binding protein 1	CG8801	FBgn0028473	CG8801	3	10	4,9015E-29	6,48E+01
	CG7927	FBgn0027549	CG7927	1	9	9,7244E-22	6,48E+01
Copia protein	copia\GIP	FBgn0013437	-	4	4	4,8911E-17	6,47E+01
	bl	FBgn0000193		10	6	4,0393E-23	6,41E+01
	CG10802	FBgn0029664	CG10802	1	5	9,7147E-27	6,41E+01
	CG5728	FBgn0039182	CG5728	1	16	1,6156E-56	6,36E+01
	Cct5	FBgn0010621	CG8439	3	12	1,964E-36	6,32E+01
	mor	FBgn0002783	CG18740	3	12	1,2364E-62	6,30E+01
Uncharacterized protein CG4951	CG4951	FBgn0039563	CG4951	2	5	3,661E-101	6,30E+01
	CG33217	FBgn0053217	CG33217	2	8	8,7422E-20	6,25E+01
Broad-complex core protein	br	FBgn0283451	CG11491	10	5	7,8657E-87	6,25E+01
39 kDa FK506-binding nuclear protein	FK506-bp1	FBgn0013269	CG6226	1	6	1,0072E-25	6,24E+01
Regucalcin	regucalcin	FBgn0030362	CG1803	3	12	1,673E-46	6,21E+01
	CG11723	FBgn0031391	CG11723	1	4	1,2163E-13	6,13E+01

	CG13887	FBgn0035165	CG13887	1	4	9,0423E-10	6,10E+01
60S ribosomal protein L5	RpL5	FBgn0064225	CG17489	4	8	7,4577E-20	6,08E+01
	Nap1	FBgn0015268	CG5330	2	5	1,8767E-25	6,05E+01
	CG4239	FBgn0030745	CG4239	3	2	5,3651E-14	6,04E+01
	pzg	FBgn0259785	CG7752	1	13	1,185E-122	5,92E+01
Heat shock protein 27	Hsp27	FBgn0001226	CG4466	1	4	2,2022E-14	5,88E+01
	Rrp6	FBgn0038269	CG7292	3	15	4,5928E-50	5,86E+01
	CG9797	FBgn0037621	CG9797	1	5	5,1533E-32	5,84E+01
RNA-binding protein 4F	Rnp4F	FBgn0014024	CG3312	4	11	5,412E-149	5,82E+01
Eukaryotic translation initiation factor 3 subunit E	Int6	FBgn0025582	CG9677	2	6	1,16E-26	5,80E+01
Caspase-8	Dredd	FBgn0020381	CG7486	4	7	3,835E-108	5,80E+01
	CG5589	FBgn0036754	CG5589	3	7	6,8411E-50	5,78E+01
	nsl1	FBgn0262527	CG4699	9	7	3,3417E-64	5,77E+01
Nuclear factor NF-kappa-B p110 subunit	Rel	FBgn0014018	CG11992	2	13	6,4436E-29	5,77E+01
	IntS4	FBgn0026679	CG12113	1	10	3,3139E-33	5,75E+01
	Aats-ile	FBgn0027086	CG11471	1	23	8,98E-100	5,70E+01
60S acidic ribosomal protein P1	RpLP1	FBgn0002593	CG4087	1	3	1,6872E-52	5,70E+01
	CG12768	FBgn0037206	CG12768	2	5	4,6713E-13	5,69E+01
	CG13895	FBgn0035158	CG13895	2	3	1,344E-109	5,68E+01
RNA-binding protein 8A	tsu	FBgn0033378	CG8781	2	5	3,4688E-14	5,66E+01
	CG4729	FBgn0036623	CG4729	3	6	1,0821E-13	5,65E+01
	Bub3	FBgn0025457	CG7581	4	5	8,7179E-25	5,62E+01
40S ribosomal protein S17	RpS17	FBgn0005533	CG3922	1	6	6,4722E-17	5,61E+01
Negative elongation factor A	Nelf-A	FBgn0038872	CG5874	1	10	9,5939E-67	5,55E+01
Pyruvate kinase	PyK	FBgn0267385	CG7070	2	15	1,376E-100	5,53E+01
	NTPase	FBgn0024947	CG3059	4	9	2,4818E-53	5,53E+01
Heat shock factor protein	Hsf	FBgn0001222	CG5748	11	12	2,0765E-88	5,52E+01
	Cortactin	FBgn0025865	CG3637	2	28	4,166E-245	5,48E+01
Nuclear pore complex protein Nup160 homolog	Nup160	FBgn0262647	CG4738	1	9	1,2611E-58	5,48E+01
	CG1888	FBgn0033421	CG1888	1	8	1,9312E-28	5,41E+01

	D19B	FBgn0022699	CG10270	2	12	2,3551E-75	5,34E+01
	shrb	FBgn0086656	CG8055	1	5	5,5149E-27	5,33E+01
	CG14442	FBgn0029893	CG14442	3	8	2,7195E-34	5,25E+01
THO complex subunit 6	thoc6	FBgn0036263	CG5632	1	5	3,6359E-15	5,19E+01
Kinesin light chain	Klc	FBgn0010235	CG5433	1	8	1,6176E-23	5,17E+01
RNA-binding protein squid	sqd	FBgn0263396	CG16901	3	4	1,0296E-29	5,13E+01
	cindr	FBgn0027598	CG31012	5	20	0	5,09E+01
	Pol32	FBgn0283467	CG3975	17	6	1,144E-16	5,09E+01
	Desat1	FBgn0086687	CG5887	8	4	4,102E-100	5,08E+01
Lysyl-tRNA synthetase	Aats-lys	FBgn0027084	CG12141	3	17	1,1773E-47	5,04E+01
	CG7824	FBgn0039711	CG7824	4	4	2,8581E-23	5,02E+01
	Sec63	FBgn0035771	CG8583	1	9	1,2809E-21	4,97E+01
Actin-57B	Act57B	FBgn0000044	CG10067	11	15	6,803E-213	4,97E+01
	bocks	FBgn0037719	CG9424	2	5	7,6709E-18	4,92E+01
	Fmr1	FBgn0028734	CG6203	6	12	3,876E-141	4,91E+01
	CG18178	FBgn0036035	CG18178	1	5	2,3493E-28	4,89E+01
	CG9799	FBgn0038146	CG9799	2	10	2,3812E-26	4,88E+01
	CG5554	FBgn0034914	CG5554	1	4	6,3678E-09	4,87E+01
Ribosome biogenesis protein BRX1 homolog	CG11583	FBgn0035524	CG11583	1	7	1,6633E-56	4,87E+01
	CG7696	FBgn0051224	CG31224	3	10	2,954E-71	4,86E+01
	CG15100	FBgn0034401	CG15100	2	11	8,1929E-71	4,85E+01
	CG1371	FBgn0033482	CG1371	5	19	1,9799E-65	4,80E+01
Nucleosomal histone kinase 1	ball	FBgn0027889	CG6386	2	5	1,1675E-10	4,80E+01
Guanine nucleotide-binding protein-like 3 homolog	Ns1	FBgn0038473	CG3983	2	9	9,8367E-37	4,78E+01
Probable elongation factor 1-delta	eEF1delta	FBgn0032198	CG4912	2	8	1,699E-102	4,74E+01
RuvB-like helicase 2	rept	FBgn0040075	CG9750	1	7	1,8192E-24	4,73E+01
	Zif	FBgn0037446	CG10267	1	6	1,5933E-22	4,73E+01
Embryonic polarity protein dorsal	dl	FBgn0260632	CG6667	4	8	3,3726E-28	4,71E+01
	CG8771	FBgn0033766	CG8771	2	9	8,851E-37	4,71E+01
60S ribosomal protein L28	RpL28	FBgn0035422	CG12740	4	3	7,4883E-09	4,67E+01

	Cnx99A	FBgn0015622	CG11958	9	11	1,8558E-36	4,65E+01
	RpL23A	FBgn0026372	CG7977	5	6	2,6935E-25	4,62E+01
Transitional endoplasmic reticulum ATPase TER94	TER94	FBgn0261014	CG2331	5	33	5,072E-247	4,61E+01
	CG17337	FBgn0259979	CG17337	2	6	1,2089E-13	4,57E+01
Ran GTPase-activating protein	RanGAP	FBgn0003346	CG9999	5	6	1,8637E-24	4,57E+01
Ubiquitin-60S ribosomal protein L40;	Ubi-p63E	FBgn0003943	CG11624	20	6	2,609E-138	4,56E+01
	Zn72D	FBgn0263603	CG5215	4	10	6,0223E-51	4,56E+01
	CG10107	FBgn0035713	CG10107	2	7	3,8274E-78	4,50E+01
Casein kinase II subunit alpha	CkIIalpha	FBgn0264492	CG17520	2	6	1,4964E-39	4,49E+01
	CG8939	FBgn0030720	CG8939	1	6	6,7406E-50	4,49E+01
	CG14712	FBgn0037924	CG14712	1	13	1,9497E-52	4,43E+01
	CG5168	FBgn0032246	CG5168	1	4	1,2957E-14	4,43E+01
Methionine aminopeptidase	und	FBgn0025117	CG4008	2	3	1,6816E-07	4,40E+01
Box A-binding factor	srp	FBgn0003507	CG3992	5	3	2,1392E-26	4,38E+01
Cysteine and histidine-rich domain-containing protein	CHORD	FBgn0029503	CG6198	1	5	1,3584E-31	4,36E+01
Host cell factor;HCF N-terminal chain;HCF C-terminal chain	Hcf	FBgn0039904	CG1710	3	29	0	4,35E+01
Eukaryotic translation initiation factor 2 subunit 2	eIF-2beta	FBgn0004926	CG4153	2	5	5,2308E-16	4,31E+01
	CG10333	FBgn0032690	CG10333	2	6	5,0023E-30	4,28E+01
ATPase ASNA1 homolog	CG1598	FBgn0033191	CG1598	2	4	2,8224E-48	4,27E+01
Protein purity of essence	poe	FBgn0011230	CG14472	2	25	6,5749E-55	4,25E+01
Protein peanut	pnut	FBgn0013726	CG8705	1	4	3,0153E-17	4,25E+01
	BEAF-32	FBgn0015602	CG10159	3	4	7,53E-13	4,24E+01
Nucleoporin GLE1	CG14749	FBgn0033316	CG14749	1	11	1,2738E-34	4,20E+01
	pch2	FBgn0051453	CG31453	1	5	3,5016E-25	4,19E+01
	deltaCOP	FBgn0028969	CG14813	4	6	8,9364E-23	4,13E+01
ATP-dependent RNA helicase p62	Rm62	FBgn0003261	CG10279	8	12	1,3816E-48	4,08E+01
Regulator of chromosome condensation	Rcc1	FBgn0002638	CG10480	1	9	2,164E-36	4,08E+01
	Mes2	FBgn0037207	CG11100	3	4	5,1456E-23	4,08E+01
V-type proton ATPase subunit B	Vha55	FBgn0005671	CG17369	2	6	2,8671E-27	4,07E+01
	CG1647	FBgn0039602	CG1647	4	6	8,3242E-41	4,04E+01

Transcription initiation factor TFIID subunit 9	e(y)1	FBgn0000617	CG6474	2	6	1,9597E-12	4,03E+01
	CG4554	FBgn0034734	CG4554	2	11	3,0483E-32	4,02E+01
	CG5885	FBgn0025700	CG5885	1	2	4,8388E-26	3,95E+01
	mip40	FBgn0034430	CG15119	1	5	9,2009E-44	3,95E+01
	msk	FBgn0026252	CG7935	4	17	1,379E-101	3,94E+01
	CG9798	FBgn0051522	CG31522	1	2	3,2806E-17	3,93E+01
	CG3756	FBgn0031657	CG3756	1	7	2,0443E-27	3,92E+01
Probable ATP-dependent RNA helicase pitchoune	pit	FBgn0266581	CG6375	1	9	1,4075E-26	3,90E+01
Protein MON2 homolog	mon2	FBgn0031985	CG8683	1	9	8,3299E-22	3,88E+01
	iPLA2-VIA	FBgn0036053	CG6718	4	10	3,6085E-23	3,86E+01
Ribosome biogenesis protein BOP1 homolog	CG5033	FBgn0028744	CG5033	2	7	4,1864E-35	3,86E+01
	htt	FBgn0027655	CG9995	7	11	6,1514E-34	3,86E+01
Probable dynactin subunit 2	Dmn	FBgn0021825	CG8269	1	7	1,8268E-31	3,85E+01
	CG3209	FBgn0034971	CG3209	3	3	7,1359E-10	3,83E+01
Phosphoglycerate kinase	Pgk	FBgn0250906	CG3127	1	9	1,5616E-17	3,83E+01
	lost	FBgn0263594	CG14648	3	8	5,3454E-60	3,81E+01
	Rbp2	FBgn0262734	CG4429	6	6	4,8577E-12	3,76E+01
	αCOP	FBgn0025725	CG7961	2	22	1,7334E-88	3,70E+01
	Gcn5	FBgn0020388	CG4107	2	8	4,004E-24	3,67E+01
Eukaryotic translation initiation factor 3 subunit B	eIF3-S9	FBgn0034237	CG4878	2	10	6,019E-34	3,65E+01
	CG13367	FBgn0025634	CG13367	3	2	8,1586E-07	3,62E+01
Elongation factor 2	EF2	FBgn0000559	CG2238	3	32	7,305E-133	3,62E+01
Ribonucleoside-diphosphate reductase subunit M2	RnrS	FBgn0011704	CG8975	1	5	1,9895E-27	3,57E+01
	mre11	FBgn0020270	CG16928	17	5	2,7424E-27	3,54E+01
	CG11138	FBgn0030400	CG11138	4	4	6,854E-10	3,53E+01
Replication protein A 70 kDa DNA-binding subunit	RpA-70	FBgn0010173	CG9633	1	11	5,765E-34	3,51E+01
Serrate RNA effector molecule homolog	Ars2	FBgn0033062	CG7843	2	13	4,0409E-31	3,49E+01
DNA damage-binding protein 1	pic	FBgn0260962	CG7769	1	8	6,3373E-21	3,49E+01
	CG10139	FBgn0033951	CG10139	1	4	4,569E-10	3,47E+01
C-terminal-binding protein	CtBP	FBgn0020496	CG7583	4	5	6,2051E-13	3,43E+01

	Tcp-1zeta	FBgn0027329	CG8231	1	18	2,272E-119	3,32E+01
Glyceraldehyde-3-phosphate dehydrogenase 1	Gapdh1	FBgn0001091	CG12055	3	4	1,3026E-11	3,31E+01
	Vps35	FBgn0034708	CG5625	4	9	1,5119E-58	3,30E+01
Lamin-B receptor	LBR	FBgn0034657	CG17952	2	10	7,7992E-51	3,27E+01
GPI mannosyltransferase 3	CG12006	FBgn0035464	CG12006	2	4	1,2456E-08	3,24E+01
Transcriptional regulator ATRX homolog	XNP	FBgn0039338	CG4548	3	8	4,2561E-25	3,24E+01
tRNA (cytosine(34)-C(5))-methyltransferase	CG6133	FBgn0026079	CG6133	1	14	2,7277E-47	3,22E+01
	gfzf	FBgn0250732	CG33546	3	5	1,0193E-20	3,22E+01
	Tcp-1eta	FBgn0037632	CG8351	1	14	3,925E-125	3,21E+01
Importin subunit alpha	Pen	FBgn0267727	CG4799	1	9	9,8293E-44	3,20E+01
	cib	FBgn0026084	CG4944	3	4	7,5372E-18	3,19E+01
WD repeat-containing protein 55 homolog	CG14722	FBgn0037943	CG14722	1	5	2,8855E-32	3,19E+01
	CG1908	FBgn0030274	CG1908	3	7	8,0181E-43	3,18E+01
Transcription factor grauzone	grau	FBgn0001133	CG33133	2	4	1,0356E-09	3,18E+01
Adenylosuccinate synthetase	CG17273	FBgn0027493	CG17273	2	4	3,8356E-10	3,18E+01
	Tudor-SN	FBgn0035121	CG7008	2	16	2,0945E-64	3,15E+01
	Strn-Mlck	FBgn0265045	CG44162	8	2	0,0017194	3,15E+01
	CG6418	FBgn0036104	CG6418	1	9	3,313E-63	3,14E+01
5-3 exoribonuclease 2 homolog	Rat1	FBgn0031868	CG10354	2	7	1,7183E-44	3,13E+01
Uncharacterized protein CG1785	CG1785	FBgn0030061	CG1785	1	5	2,9405E-45	3,12E+01
Protein hu-li tai shao	hts	FBgn0263391	CG43443	6	20	0	3,12E+01
	CG8478	FBgn0037746	CG8478	4	7	5,1732E-39	3,11E+01
Histone H2B	His2B	FBgn0001198	-	1	3	4,4606E-09	3,10E+01
	CG7668	FBgn0036929	CG7668	4	5	9,1307E-18	3,04E+01
	CG4004	FBgn0030418	CG4004	4	2	1,4009E-06	3,02E+01
	CG10465	FBgn0033017	CG10465	1	5	5,2442E-20	3,00E+01
	Rpt5	FBgn0028684	CG10370	2	7	7,5902E-25	2,96E+01
Zinc finger protein 1	zfh1	FBgn0004606	CG1322	2	5	1,9058E-69	2,96E+01
Elongation factor 1-alpha 1	Ef1alpha48D	FBgn0000556	CG8280	5	16	0	2,95E+01
Nuclear pore complex protein Nup88	mbo	FBgn0026207	CG6819	1	7	1,4685E-26	2,93E+01
Failed axon connections	fax	FBgn0014163	CG4609	6	15	1,674E-204	2,90E+01

Serine hydroxymethyltransferase	CG3011	FBgn0029823	CG3011	6	7	8,7928E-13	2,89E+01
	Fim	FBgn0024238	CG8649	7	23	0	2,89E+01
	ClC-b	FBgn0033755	CG8594	1	5	1,0191E-12	2,88E+01
	casp	FBgn0034068	CG8400	2	6	4,0346E-22	2,88E+01
Oxysterol-binding protein	Osbp	FBgn0020626	CG6708	2	8	2,7126E-29	2,87E+01
DNA polymerase alpha catalytic subunit	DNApol-alpha180	FBgn0259113	CG6349	1	7	4,3779E-52	2,87E+01
	CG4557	FBgn0029912	CG4557	2	9	1,0897E-39	2,85E+01
60S ribosomal protein L13	RpL13	FBgn0011272	CG4651	1	2	8,2245E-05	2,85E+01
	Ku80	FBgn0041627	CG18801	25	4	1,0805E-28	2,83E+01
Signal peptidase complex subunit 2	Spase25	FBgn0030306	CG1751	4	4	1,7397E-09	2,81E+01
Nucleosome-remodeling factor subunit NURF301	E(bx)	FBgn0000541	CG32346	6	14	3,1228E-48	2,81E+01
60S ribosomal protein L3	RpL3	FBgn0020910	CG4863	7	6	1,2494E-26	2,80E+01
	OstStt3	FBgn0011336	CG7748	2	3	2,0015E-05	2,78E+01
	CG12010	FBgn0035443	CG12010	2	7	7,8334E-19	2,76E+01
	CG8950	FBgn0034186	CG8950	1	5	2,6287E-15	2,74E+01
Ribosomal RNA processing protein 1 homolog	Nnp-1	FBgn0022069	CG12396	1	7	2,3744E-32	2,72E+01
	rin	FBgn0015778	CG9412	2	8	6,6106E-39	2,71E+01
	CG6370	FBgn0034277	CG6370	1	12	5,9713E-60	2,71E+01
	CG30403	FBgn0050403	CG30403	3	5	1,5528E-31	2,70E+01
Protein ultraspiracle	usp	FBgn0003964	CG4380	1	3	4,477E-06	2,69E+01
	Nop56	FBgn0038964	CG13849	3	5	6,0792E-15	2,67E+01
Coatomer subunit gamma	γCOP	FBgn0028968	CG1528	4	10	6,6728E-28	2,62E+01
	l(1)1Bi	FBgn0001341	CG6189	5	4	1,7827E-10	2,61E+01
	Pdi	FBgn0014002	CG6988	3	5	2,8491E-15	2,57E+01
Transcription initiation factor IIA subunit 1	TfIIA-L	FBgn0011289	CG5930	5	5	8,8005E-33	2,56E+01
MPN domain-containing protein CG4751	CG4751	FBgn0032348	CG4751	1	6	4,2918E-14	2,54E+01
	CG17259	FBgn0031497	CG17259	2	6	1,8888E-27	2,53E+01
Protein ref(2)P	ref(2)P	FBgn0003231	CG10360	1	7	4,3737E-34	2,53E+01
	Nup154	FBgn0021761	CG4579	5	10	9,8499E-35	2,49E+01
DNA replication licensing factor Mcm6	Mcm6	FBgn0025815	CG4039	2	7	2,6545E-29	2,48E+01

FACT complex subunit spt16	dre4	FBgn0002183	CG1828	2	13	2,7551E-54	2,46E+01
	CG3838	FBgn0032130	CG3838	2	2	2,2222E-06	2,45E+01
	Gint3	FBgn0034372	CG5469	1	4	4,1399E-07	2,44E+01
	yps	FBgn0022959	CG5654	2	4	8,9237E-44	2,42E+01
Heat shock 70 kDa protein cognate 4	Hsc70-4	FBgn0266599	CG4264	2	42	0	2,42E+01
	polybromo	FBgn0039227	CG11375	3	10	9,9483E-20	2,38E+01
	psq	FBgn0263102	CG2368	8	3	7,6262E-27	2,37E+01
Probable UDP-glucose 4-epimerase	Gale	FBgn0035147	CG12030	2	6	4,6362E-13	2,36E+01
	MBD-R2	FBgn0038016	CG10042	2	8	3,6483E-37	2,35E+01
Vinculin	Vinc	FBgn0004397	CG3299	2	13	5,5242E-47	2,34E+01
DNA-directed RNA polymerase I subunit RPA1	RpI1	FBgn0019938	CG10122	1	7	2,0374E-35	2,32E+01
	tacc	FBgn0026620	CG9765	10	8	5,3324E-25	2,31E+01
	CG12360	FBgn0283439	CG46281	2	7	1,916E-28	2,31E+01
Serine/threonine-protein phosphatase PP2A 65 kDa regulatory subunit	Pp2A-29B	FBgn0260439	CG17291	3	7	6,1621E-52	2,30E+01
	Hmu	FBgn0015737	CG3373	11	3	4,5024E-13	2,30E+01
	CG9894	FBgn0031453	CG9894	1	4	9,614E-50	2,29E+01
	CG4199	FBgn0025628	CG4199	3	8	1,944E-38	2,29E+01
Zinc finger CCCH domain-containing protein 15 homolog	CG8635	FBgn0033317	CG8635	1	3	2,0026E-08	2,29E+01
	Akap200	FBgn0027932	CG13388	3	23	0	2,27E+01
	Su(fu)	FBgn0005355	CG6054	2	5	3,4165E-15	2,27E+01
	CG9203	FBgn0267977	CG9203	2	5	2,4276E-12	2,27E+01
	E(Pc)	FBgn0000581	CG7776	3	8	2,0919E-44	2,24E+01
	l(3)mbt	FBgn0002441	CG5954	2	5	4,3159E-29	2,22E+01
DNA ligase	lig3	FBgn0038035	CG17227	2	7	3,2537E-19	2,21E+01
26S proteasome non-ATPase regulatory subunit 4	Rpn10	FBgn0015283	CG7619	1	3	5,0124E-14	2,20E+01
	CG11943	FBgn0031078	CG11943	9	16	2,0135E-38	2,19E+01
	Bap60	FBgn0025463	CG4303	1	6	4,1482E-29	2,18E+01
Dehydrogenase/reductase SDR family protein 7-like	CG7601	FBgn0027583	CG7601	1	4	1,3911E-54	2,17E+01
Caprin homolog	Capr	FBgn0042134	CG18811	2	12	1,8499E-44	2,17E+01
Protein groucho	gro	FBgn0001139	CG8384	3	5	3,6994E-18	2,16E+01

	Sec31	FBgn0033339	CG8266	4	7	1,4715E-80	2,16E+01
	CG14438	FBgn0029899	CG14438	1	8	2,8725E-19	2,13E+01
Trithorax group protein osa							
	osa	FBgn0261885	CG7467	4	7	1,251E-17	2,13E+01
Probable ubiquitin carboxyl-terminal hydrolase FAF	faf	FBgn0005632	CG1945	3	14	3,4768E-28	2,12E+01
	CG1910	FBgn0022349	CG1910	19	15	3,44E-172	2,10E+01
Helicase domino	dom	FBgn0020306	CG9696	10	9	1,9208E-28	2,10E+01
	CG6905	FBgn0265574	CG6905	1	4	5,7828E-13	2,07E+01
	Droj2	FBgn0038145	CG8863	1	2	8,9974E-06	2,07E+01
	eIF4G	FBgn0023213	CG10811	3	17	3,3454E-74	2,05E+01
	IntS8	FBgn0025830	CG5859	3	7	7,9424E-21	2,04E+01
	kis	FBgn0266557	CG3696	5	5	4,1886E-11	2,03E+01
	CG8929	FBgn0034504	CG8929	2	4	1,3481E-10	2,01E+01
	vig2	FBgn0046214	CG11844	5	4	1,1424E-10	2,01E+01
	Hsc70Cb	FBgn0026418	CG6603	2	19	9,891E-115	1,99E+01
	CG9839	FBgn0037633	CG9839	1	5	2,2452E-14	1,99E+01
	CG17129	FBgn0035151	CG17129	3	6	2,0446E-23	1,99E+01
	CG13097	FBgn0032051	CG13097	1	5	2,4625E-30	1,97E+01
Polycomb protein Scm	Scm	FBgn0003334	CG9495	1	6	6,4305E-21	1,97E+01
	swm	FBgn0002044	CG10084	3	8	7,2262E-28	1,94E+01
	CG1399	FBgn0033212	CG1399	6	9	6,8963E-31	1,94E+01
	chinmo	FBgn0086758	CG31666	3	4	1,3145E-16	1,93E+01
	CG2247	FBgn0030320	CG2247	1	3	1,5929E-06	1,93E+01
	Eb1	FBgn0027066	CG3265	6	3	2,4826E-07	1,92E+01
Calreticulin	CG9429	FBgn0005585	CG9429	3	4	3,666E-21	1,92E+01
Zinc finger protein on ecdysone puffs	Pep	FBgn0004401	CG6143	5	14	6,753E-198	1,89E+01
Histone-lysine N-methyltransferase eggless	egg	FBgn0086908	CG12196	1	5	1,0817E-23	1,88E+01
	CG9715	FBgn0036668	CG9715	3	6	6,3693E-16	1,88E+01
	CG10576	FBgn0035630	CG10576	3	7	2,3507E-21	1,88E+01
	CG8273	FBgn0037716	CG8273	3	4	2,1171E-11	1,86E+01
Chromosomal protein D1	D1	FBgn0000412	CG9745	2	3	1,9211E-08	1,85E+01

	CG42360	FBgn0259742	CG42360	2	5	2,7155E-23	1,85E+01
Nucleolar complex protein 2 homolog	CG9246	FBgn0032925	CG9246	1	4	8,3323E-35	1,85E+01
	IntS9	FBgn0036570	CG5222	1	5	1,0918E-28	1,84E+01
	CG4030	FBgn0034585	CG4030	1	2	0,00049702	1,84E+01
Actin-binding protein anillin	scra	FBgn0261385	CG2092	3	10	2,2276E-31	1,84E+01
Probable uridine-cytidine kinase	CG6364	FBgn0263398	CG6364	5	4	6,5644E-09	1,84E+01
Sodium/potassium-transporting ATPase subunit alpha	Atpα	FBgn0002921	CG5670	16	14	8,8351E-54	1,83E+01
	eIF-4B	FBgn0020660	CG10837	4	4	3,7285E-15	1,83E+01
Lethal(2)neighbour of tid protein;Lethal(2)neighbour of tid protein 2	l(2)not	FBgn0011297	CG4084	4	2	8,366E-06	1,80E+01
FAM203 family protein CG6073	CG6073	FBgn0039417	CG6073	1	4	2,1028E-23	1,79E+01
	Cenp-C	FBgn0266916	CG31258	3	8	1,5139E-17	1,79E+01
Enolase	Eno	FBgn0000579	CG17654	3	5	1,1E-21	1,78E+01
Phosphatidate cytidylyltransferase	CdsA	FBgn0010350	CG7962	4	3	1,0826E-23	1,77E+01
	Mad1	FBgn0026326	CG2072	2	6	1,8586E-13	1,76E+01
	par-6	FBgn0026192	CG5884	1	6	2,4899E-21	1,75E+01
Probable RNA-binding protein CG14230	CG14230	FBgn0031062	CG14230	1	6	3,0953E-27	1,74E+01
Transcription factor Ken	ken	FBgn0011236	CG5575	1	4	1,1542E-31	1,74E+01
	CG6838	FBgn0037182	CG6838	2	5	2,0451E-20	1,73E+01
	Mpp6	FBgn0032921	CG9250	1	3	6,3969E-16	1,73E+01
Serine/threonine-protein kinase polo	polo	FBgn0003124	CG12306	3	5	1,5782E-12	1,73E+01
Exportin-1	emb	FBgn0020497	CG13387	2	7	1,4127E-17	1,72E+01
Probable histone-binding protein Caf1	Caf1	FBgn0263979	CG4236	3	3	1,8796E-13	1,72E+01
	BubR1	FBgn0263855	CG7838	3	6	4,0908E-22	1,70E+01
	CG8545	FBgn0033741	CG8545	2	6	4,2866E-16	1,67E+01
40S ribosomal protein S3a	RpS3A	FBgn0017545	CG2168	6	5	1,5606E-17	1,67E+01
	CG2023	FBgn0037383	CG2023	1	5	2,4685E-17	1,67E+01
	CG7902	FBgn0004862	CG7902	2	5	2,3825E-12	1,67E+01
Pre-mRNA-processing factor 39	CG1646	FBgn0039600	CG1646	4	5	2,0287E-14	1,66E+01
	CG11875	FBgn0039301	CG11875	1	3	1,4024E-08	1,65E+01

Transcription factor HNF-4 homolog	Hnf4	FBgn0004914	CG9310	7	5	5,5063E-22	1,65E+01
	CG31156	FBgn0051156	CG31156	1	3	5,5514E-07	1,65E+01
Formin-like protein CG32138	CG32138	FBgn0267795	CG32138	5	7	4,1687E-16	1,63E+01
	CG12129	FBgn0033475	CG12129	1	3	1,0674E-06	1,63E+01
	mip130	FBgn0023509	CG3480	2	6	4,0441E-20	1,62E+01
	Girdin	FBgn0035411	CG12734	2	5	3,3534E-12	1,61E+01
	CG2065	FBgn0033204	CG2065	1	5	2,6533E-25	1,60E+01
Molybdenum cofactor synthesis protein cinnamon	cin	FBgn0000316	CG2945	5	3	3,1747E-10	1,60E+01
	tex	FBgn0037569	CG9615	2	3	1,4961E-13	1,60E+01
60S ribosomal protein L15;Ribosomal protein L15	RpL15	FBgn0028697	CG17420	2	2	1,08E-12	1,59E+01
	Psa	FBgn0261243	CG1009	7	6	1,1705E-12	1,58E+01
	XRCC1	FBgn0026751	CG4208	2	4	1,3203E-09	1,57E+01
Probable small nuclear ribonucleoprotein Sm D2	SmD2	FBgn0261789	CG1249	1	2	1,69E-26	1,55E+01
	Strica	FBgn0033051	CG7863	1	5	4,0753E-15	1,55E+01
Protein lin-54 homolog	mip120	FBgn0033846	CG6061	5	6	2,1873E-19	1,54E+01
AP-2 complex subunit alpha	alpha-Adaptin	FBgn0264855	CG4260	3	7	5,494E-18	1,53E+01
Protein strawberry notch	sno	FBgn0265630	CG44436	4	7	2,5732E-20	1,52E+01
	p47	FBgn0033179	CG11139	1	4	3,8118E-33	1,52E+01
	γSnap1	FBgn0028552	CG3988	3	3	4,5751E-06	1,51E+01
Ubiquitin-like modifier-activating enzyme 5	CG1749	FBgn0030305	CG1749	1	3	5,1342E-07	1,48E+01
	CG13625	FBgn0039210	CG13625	1	3	3,9093E-25	1,47E+01
Bystin	bys	FBgn0010292	CG1430	1	5	1,5452E-12	1,47E+01
	Cand1	FBgn0027568	CG5366	4	9	3,5418E-18	1,46E+01
E3 ubiquitin-protein ligase Bre1	Bre1	FBgn0086694	CG10542	1	7	1,1692E-17	1,46E+01
	CG10907	FBgn0036207	CG10907	1	3	7,5772E-08	1,45E+01
	Cul-4	FBgn0033260	CG8711	2	5	1,9626E-10	1,45E+01
	CG3995	FBgn0038472	CG3995	2	3	2,4679E-07	1,44E+01
Adenosylhomocysteinase	Ahcy13	FBgn0014455	CG11654	1	2	7,9698E-06	1,42E+01
	defl	FBgn0036038	CG18176	2	5	1,5043E-14	1,42E+01
	Karybeta3	FBgn0087013	CG1059	3	7	1,6487E-30	1,41E+01
Bifunctional methylenetetrahydrofolate	Nmdmc	FBgn0010222	CG18466	2	3	1,8966E-08	1,40E+01

dehydrogenase/cyclohydrolase, mitochondrial							
Coatomer subunit beta	betaCop	FBgn0008635	CG6223	1	9	5,7741E-25	1,39E+01
	CG3229	FBgn0053123	CG33123	2	7	2,5837E-16	1,39E+01
DNA-binding protein modulo	mod	FBgn0002780	CG2050	2	17	8,947E-131	1,39E+01
	CG14805	FBgn0023514	CG14805	3	2	4,4112E-05	1,39E+01
	CG6962	FBgn0037958	CG6962	1	4	6,8214E-46	1,39E+01
Hepatocyte growth factor-regulated tyrosine kinase substrate	Hrs	FBgn0031450	CG2903	5	5	1,6083E-27	1,37E+01
	Rpt3	FBgn0028686	CG16916	4	2	2,0464E-07	1,37E+01
	RpS15	FBgn0034138	CG8332	2	3	8,5689E-14	1,37E+01
	Nc73EF	FBgn0010352	CG11661	4	6	2,2403E-23	1,37E+01
	Srp72	FBgn0038810	CG5434	3	4	5,5323E-18	1,36E+01
	Sc2	FBgn0035471	CG10849	3	3	7,0758E-09	1,34E+01
	CG8108	FBgn0027567	CG8108	2	5	7,2578E-15	1,34E+01
Putative mitochondrial inner membrane protein	CG6455	FBgn0019960	CG6455	1	2	2,9195E-07	1,34E+01
40S ribosomal protein S10b	RpS10b	FBgn0261593	CG14206	5	3	5,1871E-09	1,32E+01
	TepIV	FBgn0041180	CG10363	17	5	2,1123E-10	1,31E+01
	Dek	FBgn0026533	CG5935	6	4	2,8816E-20	1,31E+01
	CTCF	FBgn0035769	CG8591	2	4	5,2374E-10	1,30E+01
Conserved oligomeric Golgi complex subunit 2	ldlCp	FBgn0026634	CG6177	1	4	4,5601E-13	1,27E+01
	RhoGAP18B	FBgn0261461	CG42274	5	3	1,7968E-42	1,27E+01
	CG7757	FBgn0036915	CG7757	2	4	9,8889E-09	1,27E+01
	G9a	FBgn0040372	CG2995	2	9	7,7039E-21	1,26E+01
	CG1550	FBgn0033225	CG1550	1	4	4,3693E-12	1,26E+01
HEAT repeat-containing protein 1 homolog	l(2)k09022	FBgn0086451	CG10805	1	6	4,7609E-29	1,25E+01
Aprataxin and PNK-like factor	CG6171	FBgn0026737	CG6171	3	2	4,0577E-12	1,25E+01
Actin-5C	Act5C	FBgn0000042	CG4027	1	18	0	1,24E+01
Eukaryotic translation initiation factor 5A	eIF-5A	FBgn0034967	CG3186	1	2	8,4935E-05	1,24E+01
	Brf	FBgn0038499	CG31256	3	3	3,8809E-07	1,23E+01
	l(3)72Dn	FBgn0263605	CG5018	2	5	2,5151E-09	1,23E+01
Protein penguin	pen	FBgn0015527	CG1685	1	7	9,6853E-15	1,22E+01

	Aats-his	FBgn0027087	CG6335	2	3	1,3574E-06	1,21E+01
	D12	FBgn0027490	CG13400	3	5	1,4518E-15	1,21E+01
	CG14005	FBgn0031739	CG14005	1	4	9,1957E-09	1,20E+01
	CG3335	FBgn0036018	CG3335	2	6	1,6122E-16	1,20E+01
	Hop	FBgn0024352	CG2720	4	11	4,7E-34	1,20E+01
	CG2051	FBgn0037376	CG2051	2	4	3,2977E-13	1,19E+01
	CG5953	FBgn0032587	CG5953	3	4	2,1903E-19	1,19E+01
40S ribosomal protein S18	RpS18	FBgn0010411	CG8900	3	3	8,4857E-08	1,19E+01
Probable 26S proteasome non-ATPase regulatory subunit 3	Rpn3	FBgn0261396	CG42641	3	4	1,3683E-06	1,18E+01
	CG8735	FBgn0033309	CG8735	1	3	4,8657E-07	1,18E+01
Protein NASP homolog	CG8223	FBgn0037624	CG8223	1	7	4,2907E-42	1,18E+01
	CG10425	FBgn0039304	CG10425	1	3	3,7475E-09	1,18E+01
G2/mitotic-specific cyclin-B	CycB	FBgn0000405	CG3510	4	2	6,0827E-05	1,18E+01
	CG12909	FBgn0033507	CG12909	5	3	8,6074E-09	1,17E+01
Mediator of RNA polymerase II transcription subunit 1	MED1	FBgn0037109	CG7162	5	4	1,0682E-42	1,17E+01
Heat shock protein 68	Hsp68	FBgn0001230	CG5436	1	9	1,563E-135	1,17E+01
Protein arginine N-methyltransferase 5	csul	FBgn0015925	CG3730	3	4	2,013E-10	1,16E+01
	CG12702	FBgn0031070	CG12702	2	5	4,3392E-25	1,16E+01
	CG8209	FBgn0035830	CG8209	2	4	2,2384E-10	1,15E+01
Tubulin beta-3 chain	betaTub60D	FBgn0003888	CG3401	1	8	5,6514E-38	1,14E+01
Probable tRNA (guanine(26)-N(2))-dimethyltransferase	CG6388	FBgn0032430	CG6388	1	5	1,1346E-12	1,13E+01
	CG12065	FBgn0030052	CG12065	2	10	9,0732E-79	1,13E+01
Structural maintenance of chromosomes protein	SMC2	FBgn0027783	CG10212	3	7	1,5989E-19	1,12E+01
	PMCA	FBgn0259214	CG42314	8	7	5,9153E-24	1,12E+01
	CG6841	FBgn0036828	CG6841	2	5	1,1744E-17	1,11E+01
	Patr-1	FBgn0266053	CG5208	3	5	8,4314E-29	1,11E+01
Inositol-3-phosphate synthase	Inos	FBgn0025885	CG11143	1	9	1,826E-34	1,11E+01
Tyrosine-protein kinase PR2	Ack-like	FBgn0263998	CG43741	2	4	1,6984E-11	1,10E+01
	EndoGI	FBgn0028515	CG4930	1	4	4,8847E-14	1,10E+01
	Nup50	FBgn0033264	CG2158	1	10	8,471E-40	1,09E+01
14-3-3 protein epsilon	14-3-3epsilon	FBgn0020238	CG31196	6	9	3,7559E-24	1,09E+01

Heat shock protein 26	Hsp26	FBgn0001225	CG4183	1	2	0,00022892	1,09E+01
Histone H1	CG33801	FBgn0053801	CG33801	7	3	4,5395E-20	1,09E+01
Probable elongator complex protein 2	Elp2	FBgn0033540	CG11887	1	5	1,9324E-14	1,09E+01
	mxc	FBgn0260789	CG12124	2	3	3,6173E-06	1,08E+01
	CG7556	FBgn0030990	CG7556	1	4	7,4328E-13	1,08E+01
	CG12576	FBgn0031190	CG12576	2	3	7,5219E-13	1,08E+01
Transcription initiation factor TFIID subunit 1	Taf1	FBgn0010355	CG17603	5	5	1,0866E-09	1,08E+01
	CG14696	FBgn0037853	CG14696	3	3	1,628E-07	1,07E+01
	slik	FBgn0035001	CG4527	8	10	4,941E-31	1,07E+01
Transcription factor Dp	Dp	FBgn0011763	CG4654	3	3	3,6236E-10	1,07E+01
Probable transaldolase;Transaldolase	tal	FBgn0087003	-	2	2	0,0012336	1,06E+01
	CG2260	FBgn0030000	CG2260	1	2	1,6677E-10	1,06E+01
	CG9727	FBgn0037445	CG9727	1	3	8,7966E-08	1,06E+01
	jigr1	FBgn0039350	CG17383	1	3	1,2166E-08	1,06E+01
	Graf	FBgn0030685	CG8948	2	3	8,562E-07	1,05E+01
	kcc	FBgn0261794	CG5594	4	4	6,9651E-08	1,05E+01
Oxysterol-binding protein	CG1513	FBgn0033463	CG1513	4	3	3,3063E-05	1,04E+01
	CG10289	FBgn0035688	CG10289	2	9	3,0747E-28	1,03E+01
	Aats-trp	FBgn0010803	CG9735	3	4	3,1472E-10	1,03E+01
	nito	FBgn0027548	CG2910	2	4	3,7003E-15	1,02E+01
Mediator of RNA polymerase II transcription subunit 13	skd	FBgn0003415	CG9936	3	4	9,7255E-21	1,02E+01
	kst	FBgn0004167	CG12008	8	38	1,154E-108	1,02E+01
Guanine nucleotide-binding protein subunit beta-like protein	Rack1	FBgn0020618	CG7111	1	2	0,0001196	1,02E+01
Tubulin alpha-3 chain;	alphaTub84D	FBgn0003885	CG2512	15	13	2,88E-118	1,01E+01
Protein asunder	asun	FBgn0020407	CG6814	1	2	0,000529	1,00E+01
Signal recognition particle receptor subunit alpha homolog	Gtp-bp	FBgn0010391	CG2522	2	3	1,6557E-05	1,00E+01
Protein enabled	ena	FBgn0000578	CG15112	4	3	9,782E-07	1,00E+01
	HP5	FBgn0030301	CG1745	3	2	1,0396E-05	9,99E+00
LIN1-like protein	CG5198	FBgn0032250	CG5198	2	3	2,3194E-06	9,91E+00
	nocte	FBgn0261710	CG17255	2	19	1,63E-164	9,88E+00

COP9 signalosome complex subunit 4	CSN4	FBgn0027054	CG8725	2	3	1,8513E-19	9,85E+00
	CG11883	FBgn0033538	CG11883	3	4	7,6784E-20	9,85E+00
Nucleolar protein 14 homolog	l(3)07882	FBgn0010926	CG5824	1	5	6,1635E-11	9,84E+00
	CG9153	FBgn0035207	CG9153	3	4	1,1647E-07	9,83E+00
Synaptobrevin	Syb	FBgn0003660	CG12210	5	2	3,5783E-11	9,79E+00
	Sec22	FBgn0260855	CG7359	1	2	1,3734E-12	9,77E+00
	Nopp140	FBgn0037137	CG7421	1	6	5,1408E-29	9,75E+00
	dalao	FBgn0030093	CG7055	2	2	5,6266E-09	9,71E+00
	Nat1	FBgn0031020	CG12202	1	5	2,2006E-20	9,66E+00
Heat shock protein 83	Hsp83	FBgn0001233	CG1242	1	39	0	9,63E+00
	CG9684	FBgn0037583	CG9684	2	5	1,1596E-13	9,55E+00
Poly [ADP-ribose] polymerase	Parp	FBgn0010247	CG40411	3	4	1,8115E-11	9,54E+00
	CG34422	FBgn0085451	CG34422	2	4	1,0603E-28	9,45E+00
40S ribosomal protein S4	RpS4	FBgn0011284	CG11276	1	2	0,00017634	9,41E+00
Probable N6-adenosine-methyltransferase MT-A70-like protein	CG5933	FBgn0039139	CG5933	1	4	1,8564E-11	9,36E+00
Cytosolic Fe-S cluster assembly factor NUBP1 homolog	CG17904	FBgn0032597	CG17904	1	2	9,398E-09	9,33E+00
AP-3 complex subunit delta	g	FBgn0001087	CG10986	4	3	4,559E-07	9,31E+00
	wdb	FBgn0027492	CG5643	1	5	8,1968E-12	9,31E+00
	ens	FBgn0264693	CG14998	8	5	6,4857E-20	9,29E+00
Regulator of telomere elongation helicase 1 homolog	CG4078	FBgn0029798	CG4078	4	5	1,4287E-13	9,29E+00
	CG7956	FBgn0038890	CG7956	6	3	6,1782E-07	9,22E+00
Uncharacterized protein CG7065	CG7065	FBgn0030091	CG7065	1	5	3,9676E-17	9,21E+00
	Oga	FBgn0038870	CG5871	2	6	1,0544E-13	9,14E+00
Dynamin	shi	FBgn0003392	CG18102	7	3	7,2388E-18	9,09E+00
	CG7845	FBgn0033059	CG7845	1	2	2,4103E-07	9,06E+00
Rho GTPase-activating protein 92B	RhoGAP92B	FBgn0038747	CG4755	2	6	2,2969E-13	9,05E+00
	CG6195	FBgn0038723	CG6195	1	2	0,00048434	8,90E+00
DNA-directed RNA polymerase II subunit RPB1	RpII215	FBgn0003277	CG1554	6	7	3,0028E-21	8,85E+00
Peptidyl-prolyl cis-trans isomerase	Cyp1	FBgn0004432	CG9916	3	2	5,2118E-14	8,80E+00
	CG4282	FBgn0034114	CG4282	1	3	7,4118E-19	8,77E+00
Acyl-CoA synthetase family member 4 homolog	U26	FBgn0027780	CG13401	2	3	1,4921E-07	8,76E+00

	vlc	FBgn0259978	CG8390	3	3	1,1748E-05	8,66E+00
	MEP-1	FBgn0035357	CG1244	3	4	4,0407E-14	8,66E+00
	mr	FBgn0002791	CG3060	2	5	2,0388E-22	8,64E+00
	CG15107	FBgn0041702	CG15107	1	2	5,0835E-05	8,60E+00
Molybdopterin synthase catalytic subunit	Mocs2	FBgn0039280	CG10238	1	2	8,6736E-10	8,53E+00
	Arc1	FBgn0033926	CG12505	2	2	1,6828E-19	8,49E+00
	coro	FBgn0265935	CG9446	1	4	6,128E-09	8,45E+00
Polycomb protein l(1)G0020	l(1)G0020	FBgn0027330	CG1994	2	3	2,2192E-07	8,31E+00
	CG10565	FBgn0037051	CG10565	1	4	2,635E-20	8,26E+00
	Vps20	FBgn0034744	CG4071	2	3	9,386E-08	8,18E+00
	CG9641	FBgn0031483	CG9641	2	2	5,6191E-05	8,08E+00
14-3-3 protein zeta	14-3-3zeta	FBgn0004907	CG17870	4	7	1,294E-61	8,07E+00
	CG14814	FBgn0023515	CG14814	5	2	1,0112E-06	7,94E+00
Apoptosis inhibitor 5 homolog	Aac11	FBgn0027885	CG6582	1	3	7,8433E-13	7,93E+00
Heterogeneous nuclear ribonucleoprotein 87F	Hrb87F	FBgn0004237	CG12749	3	3	3,797E-15	7,91E+00
40S ribosomal protein S14	RpS14a	FBgn0004403	CG1524	1	2	0,0000459	7,84E+00
	Ranbp9	FBgn0037894	CG5252	2	3	3,9421E-08	7,82E+00
	CG12042	FBgn0033206	CG12042	1	2	7,7458E-05	7,80E+00
Nucleolar protein 6	Mat89Ba	FBgn0261286	CG12785	1	4	4,2974E-08	7,78E+00
	egl	FBgn0000562	CG4051	3	4	7,4028E-16	7,77E+00
Probable actin-related protein 2/3 complex subunit 2	Arpc2	FBgn0032859	CG10954	1	5	6,4548E-10	7,73E+00
Putative oxidoreductase GLYR1 homolog	CG4747	FBgn0043456	CG4747	1	7	3,3759E-31	7,68E+00
Kinesin-like protein subito	sub	FBgn0003545	CG12298	25	3	7,6579E-08	7,57E+00
	CG14480	FBgn0034242	CG14480	1	3	1,5388E-08	7,56E+00
DNA-directed RNA polymerase I subunit RPA2	RpI135	FBgn0003278	CG4033	1	5	4,9236E-24	7,53E+00
	Dg	FBgn0034072	CG18250	5	2	8,4586E-05	7,52E+00
	Aats-asn	FBgn0086443	CG10687	1	3	4,4768E-07	7,52E+00
U3 small nucleolar RNA-associated protein 18 homolog	wcd	FBgn0262560	CG7989	1	4	7,0096E-11	7,51E+00
Pescadillo homolog	CG4364	FBgn0032138	CG4364	1	2	5,1574E-05	7,49E+00
Centrosomin	cnn	FBgn0013765	CG4832	6	7	9,5739E-20	7,48E+00

Polycomb protein Sfmbt	Sfmbt	FBgn0032475	CG16975	3	3	4,0994E-07	7,48E+00
	Tap42	FBgn0051852	CG31852	1	3	4,1915E-06	7,42E+00
	CG9330	FBgn0036888	CG9330	2	10	1,1122E-21	7,41E+00
Large subunit GTPase 1 homolog	Ns3	FBgn0266284	CG14788	1	3	2,7396E-08	7,39E+00
Tubulin beta-1 chain	betaTub56D	FBgn0003887	CG9277	10	11	1,0635E-56	7,37E+00
	Klp3A	FBgn0011606	CG8590	1	2	0,00041233	7,37E+00
	KP78a	FBgn0026064	CG6715	2	2	8,8687E-10	7,35E+00
	Gp93	FBgn0039562	CG5520	2	11	1,9047E-29	7,35E+00
	vig	FBgn0024183	CG4170	1	2	1,0987E-20	7,30E+00
CLIP-associating protein	chb	FBgn0021760	CG32435	1	6	1,1051E-13	7,29E+00
	CG3605	FBgn0031493	CG3605	1	2	2,0635E-05	7,23E+00
205 kDa microtubule-associated protein	Map205	FBgn0002645	CG1483	4	32	0	7,23E+00
	Trax	FBgn0038327	CG5063	3	2	0,00050649	7,18E+00
Signal recognition particle 68 kDa protein	Srp68	FBgn0035947	CG5064	1	2	0,00022222	7,07E+00
	CG4936	FBgn0038768	CG4936	1	2	2,2786E-09	7,06E+00
Protein rigor mortis	rig	FBgn0250850	CG30149	1	4	1,0598E-07	7,04E+00
	CG4294	FBgn0034742	CG4294	2	3	3,3647E-10	7,01E+00
Histone deacetylase Rpd3;Histone deacetylase	Rpd3	FBgn0015805	CG7471	2	6	1,3073E-24	6,97E+00
	Unr	FBgn0263352	CG7015	4	2	0,00018081	6,90E+00
	Pak3	FBgn0044826	CG14895	4	4	5,6715E-29	6,90E+00
THO complex protein 7	thoc7	FBgn0035110	CG17143	2	2	2,8859E-05	6,88E+00
Probable protein phosphatase CG10417	CG10417	FBgn0033021	CG10417	2	4	1,8058E-47	6,87E+00
	Txl	FBgn0035631	CG5495	2	3	1,5733E-06	6,80E+00
Moesin/ezrin/radixin homolog 1	Moe	FBgn0011661	CG10701	7	20	2,8224E-52	6,80E+00
	Hex-A	FBgn0001186	CG3001	8	2	7,8878E-06	6,76E+00
	CG9305	FBgn0032512	CG9305	2	2	1,0447E-07	6,74E+00
Protein daughter of sevenless	dos	FBgn0016794	CG1044	3	4	3,0732E-21	6,74E+00
	Pax	FBgn0041789	CG31794	9	4	3,7686E-34	6,73E+00
	CG4849	FBgn0039566	CG4849	2	4	1,3194E-10	6,67E+00
	dgt5	FBgn0033740	CG8828	1	2	1,6425E-05	6,63E+00
	CG9590	FBgn0038360	CG9590	3	3	1,7286E-07	6,58E+00

	mus201	FBgn0002887	CG10890	7	4	1,5516E-07	6,53E+00
GTP-binding protein 128up	128up	FBgn0010339	CG8340	1	2	6,2098E-07	6,51E+00
	CG12050	FBgn0032915	CG12050	2	3	1,1891E-12	6,50E+00
	CG3308	FBgn0038877	CG3308	2	3	1,6417E-08	6,47E+00
	Rcd1	FBgn0033897	CG8233	4	3	8,0595E-09	6,46E+00
	oys	FBgn0033476	CG18445	2	2	4,5583E-08	6,46E+00
Probable prefoldin subunit 2	l(3)01239	FBgn0010741	CG6302	1	3	2,3246E-08	6,46E+00
	CG7878	FBgn0037549	CG7878	1	2	1,7827E-05	6,44E+00
Inorganic pyrophosphatase	Nurf-38	FBgn0016687	CG4634	4	2	4,1714E-05	6,42E+00
	MTA1-like	FBgn0027951	CG2244	7	3	7,8562E-06	6,42E+00
	CG7275	FBgn0036500	CG7275	3	2	2,389E-16	6,40E+00
	Atac2	FBgn0032691	CG10414	1	2	2,6653E-06	6,34E+00
Phosphoglycerate mutase	Pglym78	FBgn0014869	CG1721	7	3	5,7981E-09	6,30E+00
	CG8036	FBgn0037607	CG8036	5	12	1,8533E-30	6,29E+00
Sex determination protein fruitless	fru	FBgn0004652	CG14307	14	2	1,4872E-33	6,29E+00
Cap-specific mRNA (nucleoside-2-O-)-methyltransferase 1	CG6379	FBgn0029693	CG6379	1	3	7,5101E-06	6,25E+00
Mediator of RNA polymerase II transcription subunit 6	MED6	FBgn0024330	CG9473	2	3	2,9311E-06	6,22E+00
ATP synthase subunit alpha, mitochondrial	blw	FBgn0011211	CG3612	1	12	1,2215E-38	6,16E+00
Transcription elongation factor SPT5	Spt5	FBgn0040273	CG7626	2	3	8,8986E-14	6,14E+00
	CG9754	FBgn0034617	CG9754	1	5	1,3745E-12	6,13E+00
	Hel89B	FBgn0022787	CG4261	6	5	4,6684E-11	6,11E+00
	CG14544	FBgn0039407	CG14544	1	2	6,7968E-06	6,04E+00
	CG5854	FBgn0039130	CG5854	2	2	2,0825E-06	6,02E+00
	l(2)37Cb	FBgn0086444	CG10689	1	2	0,00038795	5,93E+00
Protein 4.1 homolog	cora	FBgn0010434	CG11949	7	7	1,2578E-28	5,92E+00
	sti	FBgn0002466	CG10522	2	13	2,212E-32	5,83E+00
	tral	FBgn0041775	CG10686	1	3	6,0234E-10	5,83E+00
	CG7627	FBgn0032026	CG7627	3	6	4,2516E-29	5,82E+00
	Snx6	FBgn0032005	CG8282	1	2	8,8826E-05	5,82E+00

	eIF5B	FBgn0026259	CG10840	4	9	6,694E-99	5,82E+00
Probable cleavage and polyadenylation	Cpsf100	FBgn0027873	CG1957	2	3	1,6126E-11	5,81E+00
Probable histone-lysine N-methyltransferase Mes-4	Mes-4	FBgn0039559	CG4976	1	4	1,3245E-16	5,78E+00
Polyadenylate-binding protein	pAbp	FBgn0265297	CG5119	1	15	7,823E-99	5,76E+00
	CG5745	FBgn0038855	CG5745	1	6	2,9997E-13	5,75E+00
Protein teashirt	tsh	FBgn0003866	CG1374	2	2	2,9095E-05	5,73E+00
	CG10973	FBgn0036306	CG10973	2	4	3,9876E-08	5,72E+00
	Ns2	FBgn0034243	CG6501	1	2	0,00049013	5,71E+00
	CG6316	FBgn0052075	CG32075	1	2	8,4951E-05	5,66E+00
Protein BCL9 homolog	lgs	FBgn0039907	CG2041	1	2	3,3399E-05	5,64E+00
	Tom70	FBgn0032397	CG6756	2	2	0,00046882	5,58E+00
	nej	FBgn0261617	CG15319	2	2	0,00013938	5,54E+00
	GM130	FBgn0034697	CG11061	5	5	2,2752E-09	5,54E+00
Protein FAM188A homolog	CG7332	FBgn0030973	CG7332	1	14	1,7968E-59	5,53E+00
ATP-dependent RNA helicase bel	bel	FBgn0263231	CG9748	1	6	3,8013E-13	5,53E+00
Enhancer of mRNA-decapping protein 3	Edc3	FBgn0036735	CG6311	2	2	3,2638E-09	5,48E+00
	Phb2	FBgn0010551	CG15081	3	2	0,00012446	5,47E+00
	Cul-2	FBgn0032956	CG1512	1	5	8,9244E-19	5,42E+00
	CG1291	FBgn0035401	CG1291	2	2	4,326E-06	5,41E+00
Septin-2	Sep2	FBgn0014029	CG4173	1	3	1,7172E-14	5,41E+00
	CG17233	FBgn0036958	CG17233	4	2	0,00061821	5,39E+00
DNA replication licensing factor Mcm2	Mcm2	FBgn0014861	CG7538	1	4	3,7738E-09	5,38E+00
	wda	FBgn0039067	CG4448	1	3	1,3944E-06	5,37E+00
La-related protein	larp	FBgn0261618	CG42551	6	15	3,1112E-36	5,36E+00
	CaBP1	FBgn0025678	CG5809	1	2	0,0013801	5,36E+00
Pyruvate carboxylase	PCB	FBgn0027580	CG1516	4	51	0	5,32E+00
Protein lingerer	lig	FBgn0020279	CG8715	5	4	4,9734E-21	5,31E+00
Probable U3 small nucleolar RNA-associated protein 11	CG1789	FBgn0030063	CG1789	1	3	6,1838E-06	5,22E+00
1-acylglycerophosphocholine O-acyltransferase	CG32699	FBgn0052699	CG32699	2	3	5,2326E-07	5,20E+00
	CG13663	FBgn0039291	CG13663	2	2	0,00022037	5,19E+00

	kin17	FBgn0024887	CG5649	3	2	0,0011071	5,15E+00
	omd	FBgn0038168	CG9591	1	3	8,7505E-08	5,14E+00
Spectrin alpha chain	alpha-Spec	FBgn0250789	CG1977	1	53	7,122E-292	5,13E+00
	Saf6	FBgn0031281	CG3883	2	3	0,00001329	5,10E+00
Putative U5 small nuclear ribonucleoprotein 200 kDa helicase	CG5931	FBgn0263599	CG5931	1	10	1,6748E-23	5,07E+00
Zinc finger protein-like 1 homolog	CG5382	FBgn0038950	CG5382	2	2	8,8777E-05	5,06E+00
	Rpt4	FBgn0028685	CG3455	4	2	0,00000475	5,01E+00
	Eip63E	FBgn0005640	CG10579	9	2	4,1024E-05	5,01E+00
	Acn	FBgn0263198	CG10473	2	3	8,3235E-31	5,00E+00
Protein ROP	Rop	FBgn0004574	CG15811	1	6	4,7536E-24	4,99E+00
	IntS1	FBgn0034964	CG3173	2	2	2,4714E-05	4,96E+00
	CG30122	FBgn0050122	CG30122	6	7	1,7356E-38	4,94E+00
	CG2807	FBgn0031266	CG2807	2	5	4,7362E-13	4,92E+00
Probable prefoldin subunit 6	CG7770	FBgn0036918	CG7770	1	2	3,0458E-05	4,92E+00
26S protease regulatory subunit 8	Rpt6	FBgn0020369	CG1489	1	2	5,3352E-05	4,89E+00
	CG1309	FBgn0035519	CG1309	1	2	2,9301E-06	4,88E+00
	CG8944	FBgn0030680	CG8944	2	2	3,9391E-05	4,85E+00
GTP-binding nuclear protein Ran	Ran	FBgn0020255	CG1404	2	3	8,1544E-07	4,85E+00
Polycomb group protein Psc	Psc	FBgn0005624	CG3886	3	3	1,5289E-08	4,83E+00
DNA-binding protein Ets97D	Ets97D	FBgn0004510	CG6338	1	2	7,1028E-06	4,82E+00
	Fatp	FBgn0267828	CG46149	4	3	3,7178E-05	4,80E+00
	CG2396	FBgn0050349	CG30349	2	3	5,406E-07	4,79E+00
Membrane-associated protein Hem	Hem	FBgn0011771	CG5837	1	3	2,6826E-05	4,76E+00
	CG11120	FBgn0039250	CG11120	2	3	5,8452E-11	4,74E+00
	CG3523	FBgn0283427	CG3523	4	37	6,14E-112	4,74E+00
	CG7967	FBgn0035251	CG7967	2	2	0,00027699	4,69E+00
F-actin-capping protein subunit alpha	cpa	FBgn0034577	CG10540	1	2	1,1561E-05	4,62E+00
Myb protein	Myb	FBgn0002914	CG9045	2	2	0,00046109	4,61E+00
Putative ATP-dependent RNA helicase me31b	me31B	FBgn0004419	CG4916	2	2	7,2284E-07	4,59E+00
	sds22	FBgn0028992	CG5851	2	2	0,00033349	4,51E+00
	CG3847	FBgn0029867	CG3847	2	2	0,00030094	4,45E+00

Myosin-IB	Myo61F	FBgn0010246	CG9155	4	3	2,0889E-10	4,45E+00
	CG11208	FBgn0034488	CG11208	1	3	6,8296E-15	4,45E+00
UDP-glucose 6-dehydrogenase	sgl	FBgn0261445	CG10072	1	2	0,00033116	4,44E+00
	Snm1	FBgn0037338	CG10018	1	2	2,0713E-06	4,42E+00
	GATAd	FBgn0032223	CG5034	3	4	3,2608E-08	4,40E+00
	CG41099	FBgn0039955	CG41099	6	4	8,9337E-21	4,34E+00
Exportin-2	Cas	FBgn0022213	CG13281	1	4	1,2009E-07	4,23E+00
	uex	FBgn0262124	CG42595	1	2	0,00018537	4,22E+00
	hyx	FBgn0037657	CG11990	1	2	2,8688E-05	4,20E+00
Dipeptidyl peptidase 3	DppIII	FBgn0037580	CG7415	4	12	5,8768E-33	4,14E+00
PTB domain-containing adapter protein ced-6	ced-6	FBgn0029092	CG11804	3	2	0,00019944	4,14E+00
	Rip11	FBgn0027335	CG6606	1	2	5,0986E-23	4,13E+00
	Mms19	FBgn0037301	CG12005	2	2	7,9848E-05	4,13E+00
Actin-related protein 2	Arp2	FBgn0011742	CG9901	3	2	8,1413E-05	4,11E+00
Ubiquitin carboxyl-terminal hydrolase	CG8494	FBgn0033916	CG8494	2	2	0,00002034	4,11E+00
	Nup62	FBgn0034118	CG6251	1	2	1,2061E-07	4,11E+00
	Zpr1	FBgn0030096	CG9060	1	3	8,156E-07	4,10E+00
40S ribosomal protein S6	RpS6	FBgn0261592	CG10944	6	2	5,6836E-05	4,06E+00
Protein Mo25	Mo25	FBgn0017572	CG4083	1	2	0,00039779	4,05E+00
UPF0483 protein CG5412	CG5412	FBgn0038806	CG5412	1	4	4,4521E-11	4,03E+00

APPENDIX II. List of proteins conjugated by bioSmt3 detected *in vivo* in Drosophila melanogaster.

Protein names	Gene names	FBID KEY	ANNOTATION SYMBOL	Number of proteins	Peptides	PEP	Average intensity bioSmt3/Ctrl
Bifunctional glutamate/proline--tRNA ligase	Aats-glupro	FBgn0005674	CG5394	2	116	0	1,09E+04
14-3-3 protein zeta;14-3-3 protein epsilon	14-3-3ζ	FBgn0004907	CG17870	7	3	1,1347E-08	8,28E+02
Histone H3.3;Histone H3	His3.3A	FBgn0014857	CG5825	3	5	1,517E-28	8,09E+02
	Uba2	FBgn0029113	CG7528	1	27	0	6,92E+02
Histone H2B	His2B	FBgn0001198	-	1	2	5,4159E-19	3,50E+02
Histone H4	His4	FBgn0001200	-	1	13	1,3768E-52	3,06E+02
	lwr	FBgn0010602	CG3018	1	4	5,9244E-13	3,05E+02
ATP synthase subunit beta, mitochondrial	ATPsynβ	FBgn0010217	CG11154	3	9	4,9509E-81	2,40E+02
Heat shock 70 kDa protein cognate 4	Hsc70-4	FBgn0266599	CG4264	1	14	1,229E-188	1,59E+02
Transitional endoplasmic reticulum ATPase TER94	TER94	FBgn0261014	CG2331	4	6	4,0583E-15	1,05E+02
Tubulin beta-2 chain	βTub85D	FBgn0003889	CG9359	1	16	5,9309E-69	1,00E+02
	smt3	FBgn0264922	CG4494	1	6	3,971E-174	9,81E+01
Heat shock 70 kDa protein cognate 1	Hsc70-1	FBgn0001216	CG8937	2	7	1,3002E-72	8,09E+01
	Mdh2	FBgn0262559	CG7998	1	3	1,4716E-25	6,94E+01
Tubulin beta-1 chain	βTub56D	FBgn0003887	CG9277	6	16	9,9335E-74	6,08E+01
Heat shock protein 68	Hsp68	FBgn0001230	CG5436	7	5	2,976E-152	4,74E+01
Heat shock 70 kDa protein cognate 3	Hsc70-3	FBgn0001218	CG4147	1	4	7,2801E-25	4,57E+01
Larval serum protein 1 alpha chain	Lsp1α	FBgn0002562	CG2559	1	5	1,0979E-10	4,55E+01
40S ribosomal protein S17	RpS17	FBgn0005533	CG3922	1	3	3,1723E-22	4,45E+01
Tubulin alpha-3 chain;Tubulin alpha-1 chain	αTub84D	FBgn0003885	CG2512	3	14	7,021E-99	4,33E+01
	CG30069	FBgn0050069	CG30069	1	7	7,5049E-57	4,32E+01
ATP synthase subunit alpha, mitochondrial	blw	FBgn0011211	CG3612	1	12	7,3051E-33	3,87E+01
RNA-binding protein 8A	tsu	FBgn0033378	CG8781	1	3	2,7578E-16	3,72E+01
Actin-5C;Actin-42A	Act5C	FBgn0000042	CG4027	2	26	0	3,69E+01
Histone H2A;Histone H2A.v	CG33856	FBgn0053856	CG33856	3	2	1,4155E-05	3,43E+01
	CG4564	FBgn0263993	CG43736	3	8	1,3289E-42	3,36E+01

	CG7839	FBgn0036124	CG7839	1	3	3,5721E-40	2,94E+01
	alt	FBgn0038535	CG18212	2	4	5,7898E-36	2,71E+01
Protein FAM188A homolog	CG7332	FBgn0030973	CG7332	1	2	5,5693E-05	2,69E+01
Heat shock protein 83	Hsp83	FBgn0001233	CG1242	1	6	2,585E-100	2,68E+01
Tubulin beta-3 chain	βTub60D	FBgn0003888	CG3401	1	10	3,2615E-38	2,65E+01
	CG30403	FBgn0050403	CG30403	1	5	1,2592E-87	2,61E+01
Host cell factor	Hcf	FBgn0039904	CG1710	2	12	3,0878E-56	2,39E+01
60S ribosomal protein L13	RpL13	FBgn0011272	CG4651	1	5	9,0502E-16	2,35E+01
Pyruvate kinase	PyK	FBgn0267385	CG7070	2	3	3,2931E-35	2,34E+01
	Sec61α	FBgn0086357	CG9539	1	3	8,2835E-06	2,24E+01
Protein CDV3 homolog	CG3760	FBgn0022343	CG3760	2	3	7,2015E-13	2,15E+01
60S ribosomal protein L10a-2	RpL10Ab	FBgn0036213	CG7283	2	2	0,00093223	2,10E+01
Ubiquitin-60S ribosomal protein L40	RpL40	FBgn0003941	CG2960	5	5	9,4484E-84	1,92E+01
	CG9300	FBgn0036886	CG9300	1	6	2,4685E-23	1,79E+01
	CG8036	FBgn0037607	CG8036	2	3	2,2429E-05	1,66E+01
Isocitrate dehydrogenase [NADP]	Idh	FBgn0001248	CG7176	5	3	6,4078E-07	1,62E+01
Clathrin heavy chain	Chc	FBgn0000319	CG9012	1	4	2,5377E-07	1,61E+01
Troponin I	wupA	FBgn0004028	CG7178	11	2	5,8274E-09	1,54E+01
Tubulin alpha-2 chain	αTub85E	FBgn0003886	CG9476	1	11	1,9357E-83	1,51E+01
TBC1 domain family member CG11727	Evi5	FBgn0262740	CG11727	4	2	0,0037161	1,50E+01
Elongation factor 2	Ef2b	FBgn0000559	CG2238	3	4	1,7847E-07	1,47E+01
	Graf	FBgn0030685	CG8948	1	3	5,4938E-05	1,45E+01
	eIF4G	FBgn0023213	CG10811	2	5	8,0907E-25	1,45E+01
Calmodulin	Cam	FBgn0000253	CG8472	1	2	1,4888E-09	1,41E+01
60S ribosomal protein L14	RpL14	FBgn0017579	CG6253	1	2	2,677E-07	1,41E+01
ATP-dependent RNA helicase WM6	Hel25E	FBgn0014189	CG7269	1	2	0,00011907	1,13E+01
Glutathione S-transferase 1-1	GstD1	FBgn0001149	CG10045	1	3	9,1522E-07	1,07E+01
26S protease regulatory subunit 4	Rpt2	FBgn0015282	CG5289	1	2	1,0979E-05	1,07E+01
Titin	sls	FBgn0086906	CG1915	5	8	5,885E-23	1,05E+01
	Rtnl1	FBgn0053113	CG33113	7	3	1,0917E-07	9,65E+00
	Rab11	FBgn0015790	CG5771	1	2	0,00020589	9,36E+00

	CG31321	FBgn0051321	CG31321	1	4	2,8144E-08	8,77E+00
Nuclear cap-binding protein subunit 1	Cbp80	FBgn0022942	CG7035	2	2	0,0014937	8,76E+00
	Ku80	FBgn0041627	CG18801	1	5	1,7108E-62	8,71E+00
40S ribosomal protein S3	RpS3	FBgn0002622	CG6779	1	3	8,3258E-16	8,50E+00
	Hml	FBgn0029167	CG7002	1	2	0,00054236	7,88E+00
40S ribosomal protein SA	sta	FBgn0003517	CG14792	2	3	5,3159E-22	7,88E+00
	CG9090	FBgn0034497	CG9090	1	2	0,0014859	7,58E+00
Cofilin/actin-depolymerizing factor homolog	tsr	FBgn0011726	CG4254	1	2	4,256E-07	7,49E+00
Protein tramtrack, beta isoform	ttk	FBgn0003870	CG1856	1	2	2,017E-08	7,47E+00
Apolipophorins;Apolipophorin-2;Apolipophorin-1	Rfabg	FBgn0087002	CG11064	1	11	4,432E-139	7,46E+00
GTP-binding nuclear protein Ran	Ran	FBgn0020255	CG1404	2	2	0,0000135	7,45E+00
	CG2118	FBgn0039877	CG2118	2	6	8,0517E-32	7,42E+00
Glyceraldehyde-3-phosphate dehydrogenase 1	Gapdh1	FBgn0001091	CG12055	1	6	2,9902E-23	7,28E+00
Glyceraldehyde-3-phosphate dehydrogenase 2	Gapdh2	FBgn0001092	CG8893	2	7	2,8103E-28	7,21E+00
Actin-57B	Act57B	FBgn0000044	CG10067	6	21	1,547E-265	7,16E+00
	Mpcp	FBgn0026409	CG4994	1	5	1,3156E-08	6,69E+00
Protein hu-li tai shao	hts	FBgn0263391	CG43443	4	9	3,4994E-70	6,44E+00
	Vap-33A	FBgn0029687	CG5014	3	2	0,00015283	6,40E+00
Fructose-bisphosphate aldolase	Ald	FBgn0000064	CG6058	4	20	4,909E-191	6,32E+00
Alcohol dehydrogenase	Adh	FBgn0000055	CG3481	1	2	6,1211E-11	6,26E+00
26S proteasome non-ATPase regulatory subunit 14	Rpn11	FBgn0028694	CG18174	1	2	0,00068868	6,23E+00
	CG8552	FBgn0031990	CG8552	1	21	0	5,62E+00
Arginine kinase	Argk	FBgn0000116	CG32031	6	6	9,577E-23	5,42E+00
	regucalcin	FBgn0030362	CG1803	2	9	3,3548E-43	5,24E+00
Fat-body protein 1	Fbp1	FBgn0000639	CG17285	2	3	2,086E-11	5,04E+00
	ATPCL	FBgn0020236	CG8322	2	3	9,0326E-14	4,99E+00
Peroxiredoxin 1	Jafrac1	FBgn0040309	CG1633	1	2	2,0866E-09	4,96E+00
Protein Im not dead yet	Indy	FBgn0036816	CG3979	3	2	3,118E-06	4,91E+00
Larval serum protein 1 beta chain	Lsp1beta	FBgn0002563	CG4178	1	10	8,1206E-33	4,84E+00
	fon	FBgn0032773	CG15825	2	2	5,1244E-18	4,79E+00
Larval cuticle protein 4	Lcp4	FBgn0002535	CG2044	1	2	2,5507E-14	4,29E+00

Elongation factor 1-alpha 1	Ef1α48D	FBgn0000556	CG8280	3	13	7,4592E-46	4,15E+00
	TM9SF4	FBgn0028541	CG7364	1	2	1,7581E-06	4,05E+00
Eukaryotic translation initiation factor 3 subunit J	Adam	FBgn0027619	CG12131	1	16	0	4,05E+00
Heat shock protein 27	Hsp27	FBgn0001226	CG4466	1	2	0,0012542	4,01E+00

APPENDIX III. List of proteins conjugated by bioSUMO3 detected in HEK 293FT mammalian cells.

Protein names	Gene names	Protein IDs	Score
Ran GTPase-activating protein 1	RanGAP1	P46060	467,01
SUMO-activating enzyme subunit 2	UBA2	Q9UBT2	108,99
Small ubiquitin-related modifier 3	SUMO3	P55854	68,42
Polyubiquitin-B	UBB	P0CG47	40,54
Heat shock cognate 71 kDa protein	HSPA8	P11142	132,43
Heat shock 70 kDa protein 1A/1B	HSPA1A	P08107	154,37
DNA topoisomerase 1	TOP1	P11387	91,82
78 kDa glucose-regulated protein	HSPA5	P11021	79,57
General transcription factor II-I	GTF2I	P78347	85,58
Transcriptional regulator Kaiso	ZBTB33	Q86T24	33,57
Transcription intermediary factor 1-beta	TRIM28	Q13263	69,57
RNA-binding protein 25	RBM25	P49756	52,42
NAD-dependent deacetylase sirtuin-1	SIRT1	Q96EB6	58,17
Poly [ADP-ribose] polymerase 1	PARP1	P09874	55,94
E3 ubiquitin-protein ligase TRIM33	TRIM33	Q9UPN9	98,32
Transcription intermediary factor 1-alpha	TRIM24	O15164	60,58
DNA topoisomerase 2-alpha	TOP2A	P11388	42,20
Polyubiquitin-C	UBC	P0CG48	59,99

APPENDIX IV. List of proteins conjugated by bioUFM1 detected in HEK 293FT mammalian cells.

Protein names	Gene names	Protein IDs	Proteins	Peptides	PEP	average ratio UFM1/ctrl
Protein odr-4 homolog	ODR4	Q5SWX8	1	2	6,47E-18	3,39E+03
Phosphoribosylformylglycinamidine synthase	PFAS	O15067	1	2	2,72E-04	9,25E+02
Brain acid soluble protein 1	BASP1	P80723	1	2	1,42E-25	4,97E+02
Sorting nexin-1	SNX1	Q13596	1	2	2,69E-05	1,07E+02
La-related protein 1	LARP1	Q6PKG0	1	2	2,66E-04	7,57E+01
Band 4.1-like protein 2	EPB41L2	O43491	1	2	3,69E-06	6,25E+01
Afadin	MLLT4	P55196	1	2	4,48E-06	5,84E+01
Calmodulin-like protein 3	CALML3	P27482	1	2	8,41E-22	5,58E+01
Leucine-rich repeat-containing protein 15	LRRC15	Q8TF66	1	2	4,20E-06	5,18E+01
Cytosol aminopeptidase	LAP3	P28838	1	2	1,95E-04	3,85E+01
4F2 cell-surface antigen heavy chain	SLC3A2	P08195	1	2	6,46E-09	3,85E+01
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	Q13151	1	2	1,02E-14	3,77E+01
Cofilin-1;Cofilin-2	CFL1;CFL2	P23528	2	2	7,32E-06	3,06E+01
Kinesin-like protein KIF11	KIF11	P52732	1	2	1,14E-05	2,95E+01
Hemoglobin subunit alpha	HBA1	P69905	1	2	5,29E-06	2,89E+01
Nucleoprotein TPR	TPR	P12270	1	2	1,70E-05	2,54E+01
DNA replication licensing factor MCM3	MCM3	P25205	1	2	1,95E-05	2,54E+01
Protein FAM91A1	FAM91A1	Q658Y4	1	2	4,27E-13	1,92E+01
Nuclear migration protein nudC	NUDC	Q9Y266	1	2	6,71E-05	1,91E+01
Inosine-5-monophosphate dehydrogenase 2	IMPDH2	P12268	1	2	6,01E-04	1,48E+01
Thioredoxin-related transmembrane protein 2	TMX2	Q9Y320	1	2	2,88E-05	1,30E+01
Exportin-1	XPO1	O14980	1	2	4,07E-06	1,30E+01
Serine/arginine repetitive matrix protein 2	SRRM2	Q9UQ35	1	2	3,14E-06	1,27E+01
Vigilin	HDLBP	Q00341	1	2	6,81E-28	1,18E+01

Voltage-dependent anion-selective channel protein 3	VDAC3	Q9Y277	1	2	1,29E-117	1,08E+01
Transportin-1	TNPO1	Q92973	1	2	2,25E-06	1,03E+01
Gamma-adducin	ADD3	Q9UEY8	1	2	2,97E-06	1,03E+01
Lysozyme C	LYZ	P61626	1	2	2,84E-07	9,15E+00
ELAV-like protein 1	ELAVL1	Q15717	1	2	1,47E-06	9,15E+00
Methionine--tRNA ligase, cytoplasmic	MARS	P56192	1	2	8,52E-05	8,78E+00
Serine/threonine-protein phosphatase 4 regulatory subunit 3A	SMEK1;SMEK2	Q6IN85	2	2	7,75E-05	8,31E+00
ADP-ribosylation factor 5	ARF5	P84085	4	2	1,06E-06	8,06E+00
Phosphate carrier protein, mitochondrial	SLC25A3	Q00325	1	2	2,15E-05	7,92E+00
60S ribosomal protein L24	RPL24	P83731	1	2	1,66E-03	7,54E+00
40S ribosomal protein S16	RPS16	P62249	1	2	2,20E-05	7,32E+00
Calnexin	CANX	P27824	1	2	4,35E-07	7,31E+00
Coatomer subunit gamma-1	COPG1	Q9Y678	2	2	2,38E-05	7,25E+00
Drebrin	DBN1	Q16643	1	2	4,16E-20	6,92E+00
26S proteasome non-ATPase regulatory subunit 3	PSMD3	O43242	1	2	7,82E-08	6,87E+00
Double-stranded RNA-specific adenosine deaminase	ADAR	P55265	1	2	6,35E-05	6,78E+00
Eukaryotic translation initiation factor 5A-1-like	EIF5AL1	Q6IS14	3	2	1,28E-49	6,69E+00
Epidermal growth factor receptor substrate 15-like 1	EPS15L1	Q9UBC2	1	2	1,63E-11	6,61E+00
Proteasome subunit alpha type-7;Proteasome subunit alpha type-7-like	PSMA7;PSMA8	O14818	2	2	6,68E-06	6,42E+00
Probable ATP-dependent RNA helicase DDX10	DDX10	Q13206	1	2	2,20E-05	6,15E+00
Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	O75874	1	2	3,58E-05	6,08E+00
Treacle protein	TCOF1	Q13428	1	2	3,19E-04	6,01E+00
CAD protein;Glutamine-dependent carbamoyl-phosphate synthase;Aspartate carbamoyltransferase;Dihydroorotase	CAD	P27708	1	2	6,60E-10	5,94E+00
Nucleobindin-1	NUCB1	Q02818	1	2	3,57E-06	5,62E+00
Kinesin-like protein KIF20B	KIF20B	Q96Q89	1	2	7,52E-05	5,42E+00
Serpin H1	SERPINH1	P50454	1	2	6,64E-06	5,34E+00
Histone H1.2;Histone H1.4;Histone H1.3;Histone H1t;Histone H1.1	HIST1H1C;HIST1H1E;HIST1H1D;HIST1H1T;HIST1H1A	P16403	5	2	7,49E-06	5,23E+00
Centromere/kinetochore protein zw10 homolog	ZW10	O43264	1	2	2,78E-25	4,91E+00
Complement component 1 Q subcomponent-binding protein, mitochondrial	C1QBP	Q07021	1	2	7,14E-06	4,87E+00
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	Q14103	1	2	1,30E-07	4,86E+00
Envoplakin	EVPL	Q92817	1	2	8,03E-04	4,68E+00

Lactoylglutathione lyase	GLO1	Q04760	1	2	1,41E-04	4,65E+00
26S proteasome non-ATPase regulatory subunit 2	PSMD2	Q13200	1	2	6,73E-05	4,52E+00
40S ribosomal protein S14	RPS14	P62263	1	2	6,92E-06	4,44E+00
40S ribosomal protein S6	RPS6	P62753	1	2	3,61E-05	4,04E+00

APPENDIX V. List of proteins identified with bioID approach as possible interactors of SALL1.

Protein names	Gene names	Protein ID	Proteins	Peptides	PEP	Average Intensity SALL1/SALL1(826)
Kinesin-like protein KIF1B	KIF1B	O60333	4	2	9,5362E-05	3,88E+04
Cathepsin D	CTSD	P07339	4	7	8,129E-38	1,19E+03
Peptidyl-prolyl cis-trans isomerase FKBP5	FKBP5	Q13451	5	4	1,5793E-09	1,02E+03
Keratin, type II cytoskeletal 6B	KRT6B	P04259	2	53	0	8,66E+02
Histone H2A;Histone H2A.V	H2AFV	C9J386	2	3	2,7209E-09	6,02E+02
Dachshund homolog 1	DACH1	Q9UI36	11	12	4,415E-121	5,83E+02
Leucine-rich repeat-containing protein 15	LRRC15	Q8TF66	2	7	1,5525E-35	5,32E+02
Zinc finger protein 384	ZNF384	Q8TF68	8	8	1,497E-106	4,49E+02
Protein MCM10 homolog	MCM10	Q7L590	4	22	1,355E-175	4,38E+02
Apolipoprotein D	APOD	P05090	3	5	2,8192E-11	4,27E+02
Desmoglein-4	DSG4	Q86SJ6	2	3	1,6612E-09	4,05E+02
Ubinuclein-2	UBN2	Q6ZU65	2	12	1,9919E-47	3,77E+02
Nucleus accumbens-associated protein 1	NACC1	Q96RE7	1	11	1,911E-99	3,53E+02
Protein-glutamine gamma-glutamyltransferase E	TGM3	Q08188	2	16	1,018E-180	3,44E+02
Histone H1.4	HIST1H1E	P10412	5	5	4,5078E-11	3,14E+02
Protein Wiz	WIZ	O95785	4	9	4,9244E-38	2,87E+02
AT-rich interactive domain-containing protein 3A	ARID3A	Q99856	2	10	4,2258E-50	2,84E+02
Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	ATP2A1	O14983	4	4	7,5366E-14	2,78E+02
Skin-specific protein 32	XP32	Q5T750	1	2	0,00080305	2,76E+02
Lysine-specific demethylase 3A	KDM3A	Q9Y4C1	6	21	3,1627E-78	2,20E+02
Transducin-like enhancer protein 4	TLE4	Q04727	19	12	1,5576E-50	2,19E+02
C-terminal-binding protein 2	CTBP2	P56545	4	10	1,5047E-43	2,17E+02
Acetyl-CoA carboxylase 1;Biotin carboxylase	ACACA	Q13085	2	128	0	2,07E+02

Structural maintenance of chromosomes flexible hinge domain-containing protein 1	SMCHD1	A6NHR9	3	11	6,4485E-27	1,87E+02
Zinc finger protein castor homolog 1	CASZ1	Q86V15	2	7	1,2946E-45	1,80E+02
Transcription factor AP-2-delta	TFAP2D;TFAP2A	Q7Z6R9	2	2	1,9675E-06	1,72E+02
Corneodesmosin	CDSN	Q15517	3	3	3,3512E-21	1,61E+02
POU domain, class 2, transcription factor 1	POU2F1	P14859	22	8	1,4256E-94	1,57E+02
Ig gamma-2 chain C region	IGHG2	P01859	3	3	3,8306E-08	1,53E+02
E3 SUMO-protein ligase PIAS1	PIAS1	O75925	2	4	9,6473E-24	1,45E+02
Homeobox protein Hox-D13	HOXD13	P35453	1	8	1,1056E-52	1,43E+02
BCL-6 corepressor	BCOR	Q6W2J9	6	33	3,357E-146	1,39E+02
V-set and immunoglobulin domain-containing protein 8	VSIG8	Q5VU13	1	2	1,5847E-08	1,21E+02
40S ribosomal protein S9	RPS9	P46781	5	6	4,0826E-11	1,14E+02
ESF1 homolog	ESF1	Q9H501	1	7	1,1791E-50	1,12E+02
Protein-glutamine gamma-glutamyltransferase K	TGM1	P22735	4	11	1,9266E-30	1,10E+02
Forkhead box protein C1	FOXC1	Q12948	2	4	5,3817E-22	1,08E+02
Ig lambda-3 chain C regions;I	IGLC3	P0CG06	6	2	1,3058E-12	1,02E+02
DNA-binding protein SATB2	SATB2	Q9UPW6	3	4	8,2554E-12	1,02E+02
Friend leukemia integration 1 transcription factor	FLI1	Q01543	2	11	1,076E-132	9,78E+01
Transducin-like enhancer protein 1	TLE1	Q04724	4	12	1,4586E-46	9,60E+01
Galectin-3	LGALS3	P17931	2	3	7,0034E-20	9,29E+01
Chromosome-associated kinesin KIF4A	KIF4A	O95239	12	19	2,4676E-45	8,87E+01
Homeobox protein CDX-2	CDX2	Q99626	2	2	8,0103E-31	8,77E+01
40S ribosomal protein S28	RPS28	P62857	1	2	1,673E-08	8,28E+01
Fatty acid-binding protein, heart	FABP3	P05413	1	3	6,5531E-15	8,05E+01
Cyclin-T1	CCNT1	O60563	2	13	1,023E-167	7,92E+01
Peroxiredoxin-6	PRDX6	P30041	1	6	8,284E-15	7,84E+01
Interferon regulatory factor 2-binding protein-like	IRF2BPL	Q9H1B7	1	7	1,045E-20	7,70E+01
Plakophilin-3	PKP3	Q9Y446	2	7	5,7508E-16	7,62E+01
Thioredoxin-related transmembrane protein 1	TMX1	Q9H3N1	2	3	7,0416E-14	7,23E+01
Malate dehydrogenase, cytoplasmic	MDH1	P40925	6	3	2,6177E-08	6,42E+01
Nucleoside diphosphate kinase B	NME2	P22392	10	2	5,6259E-07	6,41E+01
Keratinocyte proline-rich protein	KPRP	Q5T749	1	4	7,8177E-09	6,29E+01

Cleavage stimulation factor subunit 3	CSTF3	Q12996	4	4	8,0477E-13	6,28E+01
Ig gamma-1 chain C region	IGHG1	P01857	1	2	3,1256E-05	6,01E+01
Caspase-14	CASP14	P31944	1	9	3,1991E-53	5,86E+01
Serpin B5	SERPINB5	P36952	4	4	7,3893E-13	5,83E+01
Cytosol aminopeptidase	LAP3	P28838	4	6	6,288E-14	5,82E+01
Leucine--tRNA ligase, cytoplasmic	LARS	Q9P2J5	7	11	1,3825E-44	5,81E+01
Cat eye syndrome critical region protein 5	CECR5	Q9BXW7	3	2	9,9269E-09	5,80E+01
F-box only protein 50	NCCRP1	Q6ZVX7	1	3	1,2251E-16	5,80E+01
40S ribosomal protein S14	RPS14	P62263	3	3	3,5785E-10	5,70E+01
Death domain-associated protein 6	DAXX	Q9UER7	5	2	7,4154E-07	5,70E+01
Pre-B-cell leukemia transcription factor 2	PBX2	P40425	11	5	3,265E-54	5,65E+01
Homeobox protein Hox-C13	HOXC13	P31276	2	2	4,3226E-09	5,21E+01
60S ribosomal protein L8	RPL8	P62917	7	2	8,1968E-07	5,20E+01
60S ribosomal protein L18	RPL18	Q07020	8	4	2,1107E-10	5,14E+01
Zymogen granule protein 16 homolog B	ZG16B	Q96DA0	1	3	1,4426E-15	5,08E+01
Crooked neck-like protein 1	CRNKL1	Q9BZJ0	4	8	3,5121E-24	5,01E+01
60S ribosomal protein L11	RPL11	P62913	4	3	2,191E-31	4,82E+01
E3 SUMO-protein ligase PIAS2	PIAS2	O75928	5	4	3,7579E-20	4,81E+01
Calmodulin-like protein 5	CALML5	Q9NZT1	1	5	3,053E-15	4,74E+01
40S ribosomal protein S7	RPS7	P62081	2	3	3,9494E-06	4,62E+01
PHD finger protein 21A	PHF21A	Q96BD5	11	10	8,495E-133	4,61E+01
40S ribosomal protein S20	RPS20	P60866	5	2	4,8928E-06	4,61E+01
Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	MCCC1	Q96RQ3	8	13	1,228E-130	4,59E+01
Transcription initiation factor TFIID subunit 6	TAF6	P49848	12	13	8,349E-99	4,56E+01
Methyltransferase-like protein 14	METTL14	Q9HCE5	2	5	1,0942E-15	4,51E+01
Adenosylhomocysteinase	AHCY	P23526	2	7	1,0809E-41	4,51E+01
Zinc finger protein 609	ZNF609	O15014	2	3	2,2067E-06	4,47E+01
THO complex subunit 4	ALYREF	Q86V81	2	2	0,00001296	4,37E+01
40S ribosomal protein S2	RPS2	P15880	7	2	9,1017E-29	4,35E+01
40S ribosomal protein S11	RPS11	P62280	1	3	1,9971E-05	4,25E+01
Copine-3	CPNE3	O75131	17	11	6,6063E-49	4,25E+01

HMG box transcription factor BBX	BBX	Q8WY36	6	13	7,4857E-45	4,19E+01
Integrator complex subunit 12	INTS12	Q96CB8	7	14	4,825E-158	4,17E+01
Myoglobin	MB	P02144	7	2	1,1818E-08	4,17E+01
N-alpha-acetyltransferase 15	NAA15	Q9BXJ9	8	7	1,7813E-13	4,10E+01
Homeobox protein Nkx-2.5	NKX2-5	P52952	4	4	4,5053E-22	4,06E+01
Homeobox protein Nkx-2.1	NKX2-1	P43699	3	4	1,9355E-32	4,02E+01
60S ribosomal protein L22	RPL22	P35268	1	2	1,174E-06	4,01E+01
Ligand-dependent corepressor	LCOR	Q96JN0	2	5	1,2389E-19	3,94E+01
Putative uncharacterized protein C11orf80	C11orf80	B4DXL1	9	2	0,0011845	3,90E+01
Dual specificity protein phosphatase 14	DUSP14	O95147	1	3	2,6351E-06	3,88E+01
Arachidonate 12-lipoxygenase, 12R-type	ALOX12B	O75342	1	3	3,1405E-06	3,87E+01
Inosine-5-monophosphate dehydrogenase 2	IMPDH2	P12268	4	7	5,033E-29	3,86E+01
Vacuolar protein sorting-associated protein VTA1 homolog	VTA1	Q9NP79	3	4	1,96E-41	3,80E+01
40S ribosomal protein S26	RPS26	P62854	3	2	1,5923E-11	3,78E+01
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	Q07666	12	5	6,9157E-16	3,77E+01
Homeobox protein Nkx-2.3	NKX2-3	Q8TAU0	1	4	8,6809E-14	3,75E+01
Serpin B4	SERPINB4	P48594	4	11	1,448E-162	3,73E+01
Prickle-like protein 3	PRICKLE3	O43900	7	6	1,7849E-20	3,73E+01
40S ribosomal protein S8	RPS8	P62241	2	4	5,2393E-22	3,65E+01
Coatomer subunit alpha;Xenin;Proxenin	COPA	P53621	2	4	2,2889E-10	3,64E+01
Zinc finger and BTB domain-containing protein 9	ZBTB9	Q96C00	1	5	1,2993E-33	3,62E+01
Transcription factor Sp1	SP1	P08047	2	2	9,886E-187	3,61E+01
Thioredoxin	TXN	P10599	2	4	1,6354E-13	3,54E+01
40S ribosomal protein S15a	RPS15A	P62244	1	2	0,0021831	3,52E+01
Transcription factor jun-B	JUNB	P17275	1	4	1,9348E-18	3,42E+01
ADP/ATP translocase 3;ADP/ATP translocase 1	SLC25A6;SLC25A4	P12236	3	9	4,4843E-30	3,36E+01
40S ribosomal protein S16	RPS16	P62249	2	3	6,3781E-10	3,25E+01
Homeobox protein cut-like 1;Protein CASP	CUX1	P39880	11	14	9,1898E-56	3,25E+01
Catalase	CAT	P04040	1	15	4,2647E-72	3,23E+01
Insulin receptor substrate 4	IRS4	O14654	1	7	7,5377E-33	3,20E+01
Collagen alpha-2(I) chain	COL1A2	P08123	2	3	5,0324E-06	3,18E+01

1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1	PLCD1	P51178	7	4	2,7128E-12	3,04E+01
GTP-binding nuclear protein Ran	RAN	P62826	5	4	3,2953E-08	3,04E+01
Proteasome subunit alpha type-3	PSMA3	P25788	4	2	0,00024178	3,03E+01
ADP-ribosylation factor 3	ARF3	P61204	15	4	7,3427E-10	3,02E+01
Translationally-controlled tumor protein	TPT1	P13693	5	3	7,0254E-07	3,02E+01
RNA-binding protein FUS	FUS	P35637	5	6	2,6746E-41	2,99E+01
Forkhead box protein P1	FOXP1	Q9H334	12	2	9,2828E-09	2,98E+01
Gamma-enolase;Enolase	ENO2	P09104	7	3	4,1445E-12	2,95E+01
60S ribosomal protein L10	RPL10	P27635	5	3	1,82E-07	2,93E+01
60S ribosomal protein L19	RPL19	P84098	1	2	0,00082641	2,91E+01
Phospholipase D3	PLD3	Q8IV08	2	2	9,7906E-06	2,80E+01
Hephaestin-like protein 1	HEPHL1	Q6MZM0	1	4	1,1443E-10	2,78E+01
60S ribosomal protein L27	RPL27	P61353	1	3	3,6023E-05	2,74E+01
Stathmin-2;Stathmin	STMN2	Q93045	4	2	1,7274E-05	2,73E+01
60S ribosomal protein L14	RPL14	P50914	3	2	1,0132E-07	2,73E+01
E3 SUMO-protein ligase PIAS4	PIAS4	Q8N2W9	1	4	5,1596E-12	2,73E+01
G patch domain-containing protein 11	GPATCH11	Q8N954	2	3	3,8929E-06	2,72E+01
Bloom syndrome protein	BLM	P54132	4	7	6,5826E-20	2,71E+01
Zinc finger and BTB domain-containing protein 21	ZBTB21	Q9ULJ3	2	13	2,874E-103	2,70E+01
Forkhead box protein J3	FOXJ3	Q9UPW0	20	2	3,8174E-06	2,69E+01
Histone H3.3	H3F3A	P84243	6	6	6,9418E-14	2,69E+01
Histone H2A.Z	H2AFZ	P0C0S5	5	3	1,6451E-08	2,68E+01
Far upstream element-binding protein 1	FUBP1	Q96AE4	7	9	1,0119E-23	2,65E+01
40S ribosomal protein S24	RPS24	P62847	5	2	1,6015E-06	2,63E+01
14-3-3 protein sigma	SFN	P31947	6	13	7,1541E-86	2,63E+01
Nuclear pore complex protein Nup155	NUP155	O75694	4	2	4,7227E-09	2,62E+01
60S ribosomal protein L7	RPL7	P18124	4	2	4,7478E-07	2,55E+01
Acidic leucine-rich nuclear phosphoprotein 32 family member A	ANP32A	P39687	7	2	3,4089E-05	2,54E+01
Secretory carrier-associated membrane protein 3	SCAMP3	O14828	2	2	1,4883E-13	2,52E+01
Homeobox protein DLX-3	DLX3	O60479	2	3	1,1737E-10	2,50E+01
40S ribosomal protein S6	RPS6	P62753	3	2	0,0016369	2,46E+01

Homeobox protein SIX4	SIX4	Q9UIU6	3	8	1,5481E-58	2,45E+01
Proteasome activator complex subunit 3	PSME3	P61289	4	3	2,9958E-21	2,44E+01
Histone H1.5	HIST1H1B	P16401	1	3	9,9713E-08	2,41E+01
60S ribosomal protein L15;Ribosomal protein L15	RPL15	P61313	4	2	3,3452E-05	2,41E+01
Triosephosphate isomerase	TPI1	P60174	3	14	1,058E-229	2,37E+01
Zinc finger protein 536	ZNF536	O15090	22	3	8,3185E-07	2,35E+01
Zinc finger protein 362	ZNF362	Q5T0B9	2	9	1,633E-135	2,33E+01
14-3-3 protein gamma;14-3-3 protein gamma, N-terminally processed	YWHAG	P61981	3	4	9,1E-27	2,33E+01
ELM2 and SANT domain-containing protein 1	ELMSAN1	Q6PJG2	2	3	2,2582E-06	2,31E+01
RNA-binding protein 6	RBM6	P78332	7	5	7,2129E-12	2,31E+01
Ig kappa chain C region	IGKC	P01834	1	3	4,8714E-79	2,28E+01
Beta-catenin-like protein 1	CTNNBL1	Q8WYA6	3	3	6,5412E-06	2,26E+01
C-terminal-binding protein 1	CTBP1	Q13363	9	6	2,814E-15	2,24E+01
Prelamin-A/C;Lamin-A/C	LMNA	P02545	10	16	2,5144E-68	2,23E+01
YY1-associated protein 1	YY1AP1	Q9H869	15	10	9,0788E-55	2,21E+01
Lysozyme C	LYZ	P61626	2	3	1,0693E-10	2,18E+01
Motor neuron and pancreas homeobox protein 1	MNX1	P50219	4	2	6,1139E-14	2,13E+01
Homeobox protein DLX-2;Homeobox protein DLX-5	DLX2;DLX5	Q07687	3	3	2,8865E-22	2,06E+01
Trifunctional enzyme subunit alpha, mitochondrial;Long-chain enoyl-CoA hydratase;Long chain 3-hydroxyacyl-CoA dehydrogenase	HADHA	P40939	2	2	0,0035944	2,02E+01
5-3 exoribonuclease 2	XRN2	Q9H0D6	3	3	8,4675E-07	1,99E+01
Polymerase delta-interacting protein 3	POLDIP3	Q9BY77	7	12	1,8262E-58	1,96E+01
Nuclear receptor subfamily 2 group C member 1	NR2C1	P13056	3	6	2,8869E-42	1,93E+01
Cip1-interacting zinc finger protein	CIZ1	Q9ULV3	10	9	9,0782E-58	1,91E+01
DNA replication licensing factor MCM7	MCM7	P33993	6	4	2,4309E-14	1,90E+01
Serpin B3	SERPINB3	P29508	2	14	5,473E-187	1,88E+01
Homeobox protein aristaless-like 4	ALX4	Q9H161	1	3	4,177E-18	1,84E+01
UPF0544 protein C5orf45	C5orf45	Q6NTE8	15	4	3,4331E-26	1,84E+01
60S ribosomal protein L12	RPL12	P30050	2	2	2,5126E-13	1,83E+01
Protein S100-A9	S100A9	P06702	1	5	1,5321E-81	1,81E+01
Protein lin-9 homolog	LIN9	Q5TKA1	5	2	0,00004325	1,80E+01
14-3-3 protein zeta/delta	YWHAZ	P63104	11	8	1,1472E-80	1,80E+01

SWI/SNF complex subunit SMARCC2	SMARCC2	Q8TAQ2	6	11	8,5267E-25	1,75E+01
NAD-dependent protein deacetylase sirtuin-1;SirtT1 75 kDa fragment	SIRT1	Q96EB6	4	8	7,4413E-91	1,74E+01
Hepatocyte nuclear factor 3-alpha	FOXA1	P55317	2	2	5,549E-11	1,73E+01
Peptidyl-prolyl cis-trans isomerase A;Peptidyl-prolyl cis-trans isomerase A, N-terminally processed;Peptidyl-prolyl cis-trans isomerase	PPIA	P62937	7	4	1,8239E-12	1,67E+01
Peptidyl-prolyl cis-trans isomerase B	PPIB	P23284	1	3	1,1596E-06	1,67E+01
Zinc finger and BTB domain-containing protein 10	ZBTB10	Q96DT7	5	3	1,8754E-12	1,65E+01
Transcription factor COE3;Transcription factor COE4;Transcription factor COE1	EBF3;EBF4;EBF1	Q9H4W6	10	3	5,0168E-16	1,64E+01
Homeobox protein Hox-A3;Homeobox protein Hox-D3	HOXA3;HOXD3	O43365	7	3	2,8687E-13	1,62E+01
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	DLAT	P10515	5	2	3,7894E-08	1,61E+01
Eukaryotic translation initiation factor 4H	EIF4H	Q15056	2	2	1,777E-08	1,60E+01
Transmembrane emp24 domain-containing protein 10	TMED10	P49755	2	2	5,4271E-05	1,59E+01
14-3-3 protein beta/alpha;14-3-3 protein beta/alpha, N-terminally processed	YWHAB	P31946	2	4	7,9399E-13	1,57E+01
28 kDa heat- and acid-stable phosphoprotein	PDAP1	Q13442	1	4	2,0898E-10	1,55E+01
DNA endonuclease RBBP8	RBBP8	Q99708	4	7	5,0603E-36	1,53E+01
Hemoglobin subunit beta	HBB	P68871	14	7	6,0915E-37	1,52E+01
60S ribosomal protein L13a;Putative 60S ribosomal protein L13a-like MGC87657	RPL13A	P40429	2	2	0,00029833	1,51E+01
Retinoic acid receptor RXR-alpha	RXRA	P19793	3	3	6,0261E-08	1,50E+01
Creatine kinase M-type;Creatine kinase M-type, N-terminally processed	CKM	P06732	1	2	2,315E-08	1,49E+01
Myocyte-specific enhancer factor 2A	MEF2A	Q02078	19	4	1,9926E-21	1,47E+01
40S ribosomal protein S15	RPS15	P62841	1	2	1,8904E-05	1,43E+01
Serine/arginine-rich splicing factor 1	SRSF1	Q07955	3	5	1,473E-14	1,43E+01
Beta-enolase;Enolase	ENO3	P13929	5	2	2,3544E-11	1,39E+01
E3 SUMO-protein ligase CBX4	CBX4	O00257	2	2	0,00017308	1,38E+01
Transcription factor jun-D	JUND	P17535	1	3	6,5308E-17	1,38E+01
60S ribosomal protein L38	RPL38	P63173	1	2	8,436E-07	1,37E+01
Complement C4-B;	C4B;C4A	P0C0L5	14	2	2,4797E-05	1,36E+01
ADP-ribosylation factor-like protein 8B	ARL8B	Q9NVJ2	2	2	1,5762E-06	1,36E+01
Protein POF1B	POF1B	Q8WVV4	5	5	2,8928E-16	1,36E+01
Gamma-glutamyl hydrolase	GGH	Q92820	2	3	2,8582E-08	1,35E+01
NADP-dependent malic enzyme	ME1	P48163	4	3	2,0766E-05	1,34E+01

CAP-Gly domain-containing linker protein 1	CLIP1	P30622	15	13	1,1497E-37	1,34E+01
Friend leukemia integration 1 transcription factor	FLI1	G3V183	12	11	8,937E-122	1,33E+01
Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	HADH	Q16836	3	2	1,8449E-05	1,33E+01
	HIST2H3PS2	Q5TEC6	1	4	2,6109E-12	1,32E+01
Xanthine dehydrogenase/oxidase;Xanthine dehydrogenase;Xanthine oxidase	XDH	P47989	1	3	1,5117E-05	1,31E+01
40S ribosomal protein S4, X isoform	RPS4X	P62701	6	6	3,0527E-14	1,27E+01
Iroquois-class homeodomain protein IRX-4	IRX4	P78413	9	2	3,4213E-08	1,26E+01
Homeobox protein PKNOX1	PKNOX1	P55347	10	2	0,0011595	1,23E+01
GA-binding protein subunit beta-1	GABPB1	Q06547	5	2	0,00001999	1,22E+01
Transducin-like enhancer protein 3	TLE3	Q04726	7	16	4,529E-151	1,22E+01
Myosin-10	MYH10	P35580	5	21	7,5155E-96	1,21E+01
Myosin-7;Myosin-6	MYH7;MYH6	P12883	6	10	5,5505E-36	1,21E+01
Tetratricopeptide repeat protein 31	TTC31	Q49AM3	6	2	0,0006191	1,17E+01
Histone H1.0;Histone H1.0, N-terminally processed	H1F0	P07305	3	3	3,8639E-07	1,16E+01
AT-rich interactive domain-containing protein 3B	ARID3B	Q8IVW6	4	41	0	1,14E+01
Leucine zipper putative tumor suppressor 2	LZTS2	Q9BRK4	2	2	6,1276E-05	1,14E+01
Small nuclear ribonucleoprotein Sm D3	SNRPD3	P62318	2	2	8,6853E-05	1,10E+01
RNA-binding protein 14	RBM14	Q96PK6	1	2	0,00024037	1,03E+01
Alpha-2-macroglobulin	A2M	P01023	2	2	2,3479E-06	1,02E+01
Doublesex- and mab-3-related transcription factor A2	DMRTA2	Q96SC8	2	2	0,0011745	1,01E+01
Protein capicua homolog	CIC	Q96RK0	1	8	6,2393E-27	9,76E+00
Desmocollin-3	DSC3	Q14574	2	5	1,9839E-19	9,69E+00
Retinoic acid-induced protein 1	RAI1	Q7Z5J4	9	17	4,0253E-48	9,63E+00
Pre-B-cell leukemia transcription factor 1	PBX1	P40424	7	2	1,1487E-09	9,60E+00
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	P49748	5	2	3,4434E-05	9,51E+00
CCAAT/enhancer-binding protein beta	CEBPB	P17676	1	3	4,5276E-10	9,43E+00
Fatty acid-binding protein, epidermal	FABP5	Q01469	1	4	1,6598E-10	9,40E+00
Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA	P67775	7	2	7,6955E-07	9,31E+00
14-3-3 protein theta	YWHAQ	P27348	3	5	1,6952E-43	9,25E+00
Small nuclear ribonucleoprotein-associated proteins B and B	SNRPB	P14678	7	2	0,001119	9,16E+00
60S acidic ribosomal protein P2	RPLP2	P05387	2	5	2,1804E-61	9,15E+00

Calponin-1	CNN1	P51911	3	2	0,00063203	9,07E+00
Small ubiquitin-related modifier 2	SUMO2	P61956	3	4	4,126E-71	8,50E+00
Methyl-CpG-binding domain protein 4	MBD4	O95243	4	2	8,1827E-06	8,49E+00
Chromatin assembly factor 1 subunit B	CHAF1B	Q13112	1	8	2,2352E-35	8,49E+00
60S ribosomal protein L31	RPL31	P62899	7	2	3,8752E-27	8,49E+00
Zinc finger protein 322	ZNF322	Q6U7Q0	2	3	1,1093E-07	8,49E+00
Thymidine phosphorylase	TYMP	P19971	2	2	0,00050131	8,25E+00
Erlin-1;Erlin-2	ERLIN1;ERLIN2	O75477	6	2	2,5675E-05	7,99E+00
Ran-specific GTPase-activating protein	RANBP1	P43487	8	6	2,79E-19	7,72E+00
PHD finger protein 3	PHF3	Q92576	9	25	1,594E-106	7,47E+00
Plakophilin-1	PKP1	Q13835	2	11	5,1108E-67	7,43E+00
Histone H4	HIST1H4A	P62805	1	10	2,0543E-38	7,17E+00
Pituitary homeobox 2	PITX2;PITX1	Q99697	8	2	1,5105E-23	7,11E+00
Aldehyde dehydrogenase, mitochondrial	ALDH2	P05091	5	4	5,2778E-09	7,05E+00
Cystatin-M	CST6	Q15828	1	2	0,00017686	7,02E+00
Regulation of nuclear pre-mRNA domain-containing protein 1A	RPRD1A	Q96P16	3	2	1,8266E-07	6,96E+00
Myosin-1;Myosin-2	MYH1;MYH2	P12882	2	4	6,9113E-09	6,95E+00
ATPase family AAA domain-containing protein 3A	ATAD3A	Q9NVI7	12	2	0,00061518	6,92E+00
Chromodomain-helicase-DNA-binding protein 3	CHD3	Q12873	3	22	7,382E-159	6,88E+00
mRNA-decapping enzyme 1B	DCP1B	Q8IZD4	4	2	0,00047938	6,86E+00
WW domain-binding protein 11	WBP11	Q9Y2W2	3	8	1,0382E-33	6,86E+00
Transcriptional enhancer factor TEF-1	TEAD1	P28347	5	2	6,8644E-06	6,62E+00
Tyrosine-protein kinase Fyn	FYN;YES1	P06241	6	2	7,657E-09	6,59E+00
Complement C3;	C3	P01024	2	2	4,0914E-06	6,57E+00
Monocarboxylate transporter 1	SLC16A1	P53985	4	2	3,1283E-07	6,52E+00
Zinc finger protein 598	ZNF598	Q86UK7	4	10	3,9382E-25	6,44E+00
Lupus La protein	SSB	P05455	6	17	5,061E-137	6,44E+00
UPF0469 protein KIAA0907	KIAA0907	Q7Z7F0	4	3	4,9039E-18	6,43E+00
Alpha-aminoadipic semialdehyde dehydrogenase	ALDH7A1	P49419	9	7	2,7999E-19	6,32E+00
Lipocalin-1	LCN1P1	P31025	2	2	6,7891E-05	6,32E+00
Protein Spindly	SPDL1	Q96EA4	12	17	4,8497E-78	6,28E+00

UPF0428 protein CXorf56	CXorf56	Q9H5V9	3	2	1,2541E-08	6,27E+00
Calpain-1 catalytic subunit	CAPN1	P07384	11	2	0,00041066	6,12E+00
TCF3 fusion partner	TFPT	P0C1Z6	3	3	1,5233E-07	6,12E+00
Transducin-like enhancer protein 3	TLE3	Q04726	17	16	1,989E-154	6,12E+00
Transcriptional regulator ATRX	ATRX	P46100	7	11	1,6223E-78	6,11E+00
Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	P41091	4	3	4,0023E-15	6,07E+00
NGFI-A-binding protein 2	NAB2	Q15742	1	2	3,8499E-05	6,04E+00
Plectin	PLEC	Q15149	11	8	9,1507E-17	6,01E+00
Transcription factor 20	TCF20	Q9UGU0	3	51	0	5,93E+00
60S ribosomal protein L3	RPL3	P39023	6	5	2,629E-11	5,92E+00
Serpin A12	SERPINA12	Q8IW75	1	2	1,9528E-05	5,90E+00
10 kDa heat shock protein, mitochondrial	HSPE1	P61604	3	3	1,0125E-10	5,88E+00
Protein FAM122A	FAM122A	Q96E09	1	2	5,7051E-05	5,88E+00
Collagen alpha-1(I) chain	COL1A1	P02452	1	3	6,5678E-07	5,84E+00
Eukaryotic translation initiation factor 3 subunit F	EIF3F	O00303	5	9	1,6671E-47	5,81E+00
Puromycin-sensitive aminopeptidase	NPEPPS	P55786	11	4	1,2497E-13	5,75E+00
Myb/SANT-like DNA-binding domain-containing protein 2	MSANTD2	Q6P1R3	5	9	5,5892E-44	5,72E+00
Hemoglobin subunit alpha	HBA1	P69905	1	3	4,4285E-11	5,71E+00
Prolyl endopeptidase	PREP	P48147	1	2	6,471E-06	5,70E+00
Protein S100-A7;Protein S100-A7A	S100A7;S100A7A	P31151	2	2	1,3658E-06	5,63E+00
Lactotransferrin	LTF	P02788	6	6	1,3446E-12	5,59E+00
40S ribosomal protein S10	RPS10	P46783	3	3	1,4909E-06	5,57E+00
Ig alpha-1 chain C region;Ig alpha-2 chain C region	IGHA1;IGHA2	P01876	2	6	1,5383E-26	5,53E+00
Zinc finger protein 830	ZNF830	Q96NB3	1	7	2,2668E-41	5,45E+00
Serine/arginine-rich splicing factor 7	SRSF7	Q16629	6	4	2,0134E-11	5,44E+00
Prothymosin alpha	PTMA	P06454	6	2	2,1519E-07	5,43E+00
Eukaryotic translation initiation factor 5A-1	EIF5A	P63241	7	2	5,0305E-06	5,40E+00
Protein FAM49B	FAM49B	Q9NUQ9	9	2	1,8104E-09	5,39E+00
Sal-like protein 1	SALL1	Q9NSC2	10	81	0	5,38E+00
Nuclear receptor subfamily 2 group C member 2	NR2C2	P49116	5	14	1,1164E-65	5,37E+00
Alpha-amylase 1	AMY1A	P04745	7	4	4,9833E-31	5,37E+00

Histone H2B type 1-B	HIST1H2BB	P33778	7	7	4,1769E-60	5,37E+00
Chromobox protein homolog 8	CBX8	Q9HC52	5	3	1,5318E-06	5,36E+00
Sororin	CDCA5	Q96FF9	3	4	3,4037E-12	5,36E+00
Casein kinase I isoform epsilon;Casein kinase I isoform delta	CSNK1E;CSNK1D	P49674	11	6	2,7268E-14	5,33E+00
Spliceosome-associated protein CWC15 homolog	CWC15	Q9P013	1	2	3,9151E-05	5,32E+00
4-trimethylaminobutyraldehyde dehydrogenase	ALDH9A1	P49189	2	3	7,3115E-09	5,25E+00
Galectin-7	LGALS7	P47929	1	3	6,9689E-15	5,22E+00
Pituitary homeobox 1	PITX1	P78337	2	2	7,8424E-24	5,22E+00
40S ribosomal protein S13	RPS13	P62277	2	2	1,0464E-05	5,21E+00
Centromere-associated protein E	CENPE	Q02224	4	2	0,00022538	5,19E+00
Zinc fingers and homeoboxes protein 3	ZHX3	Q9H4I2	6	10	3,1813E-74	5,18E+00
Homeobox protein Hox-A10	HOXA10	P31260	8	3	3,4834E-06	5,17E+00
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	P24752	4	3	1,2445E-13	5,17E+00
Tuftelin	TUFT1	F5H607	5	4	5,8048E-06	5,16E+00
Histone H2B type 1-L	HIST1H2BL	Q99880	12	7	1,1949E-60	5,12E+00
Serine/arginine-rich splicing factor 10	SRSF10	O75494	8	2	1,4772E-08	5,11E+00
Tubulin alpha-3C/D chain;Tubulin alpha-3E chain	TUBA3C;TUBA3E	Q13748	3	21	0	5,09E+00
Cyclin-dependent kinase 12	CDK12	Q9NYV4	28	9	7,3889E-30	5,04E+00
Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	P31943	24	6	4,6516E-95	5,03E+00
Insulin-degrading enzyme	IDE	P14735	3	5	1,9355E-07	5,02E+00
Protein disulfide-isomerase	P4HB	P07237	5	12	1,0765E-47	4,99E+00
Glucosidase 2 subunit beta	PRKCSH	P14314	3	2	1,1842E-06	4,96E+00
Alpha-crystallin B chain	CRYAB	P02511	5	2	0,0012907	4,96E+00
Pre-mRNA-processing factor 17	CDC40	O60508	2	4	1,3277E-11	4,95E+00
Histidine--tRNA ligase, cytoplasmic	HARS	P12081	15	4	2,6633E-10	4,91E+00
Stathmin	STMN1	P16949	3	3	4,5874E-07	4,89E+00
Peroxiredoxin-4	PRDX4	Q13162	2	2	2,5361E-07	4,87E+00
Cleavage and polyadenylation specificity factor subunit 7	CPSF7	Q8N684	16	4	3,4162E-13	4,86E+00
Cofilin-1	CFL1	P23528	9	3	3,9554E-22	4,86E+00
40S ribosomal protein S19	RPS19	P39019	1	2	2,1774E-05	4,86E+00
TIP41-like protein	TIPRL	O75663	2	6	9,5229E-13	4,85E+00

Myelin basic protein	MBP	P02686	21	3	4,5295E-21	4,84E+00
High mobility group protein B1	HMGB1	P09429	6	3	1,7706E-12	4,83E+00
Serpin B7	SERPINB7	O75635	3	2	0,00011782	4,80E+00
RB1-inducible coiled-coil protein 1	RB1CC1	Q8TDY2	2	2	0,0010934	4,79E+00
Epiplakin	EPPK1	P58107	1	6	4,6819E-28	4,74E+00
Extracellular matrix protein 1	ECM1	Q16610	4	2	8,6653E-11	4,73E+00
Homeobox protein DLX-6;Homeobox protein DLX-1	DLX6;DLX1	P56179	5	3	1,4938E-25	4,73E+00
Exosome component 10	EXOSC10	Q01780	3	26	2,942E-89	4,71E+00
60S ribosomal protein L13	RPL13	P26373	2	5	1,2011E-42	4,71E+00
Homeobox protein Hox-B8	HOXB8	P17481	6	2	1,1989E-06	4,70E+00
Histone H1x	H1FX	Q92522	1	2	3,6339E-05	4,69E+00
Profilin-1	PFN1	P07737	1	2	3,2539E-05	4,68E+00
Zinc finger protein with KRAB and SCAN domains 4	ZKSCAN4	Q969J2	2	7	8,2938E-25	4,66E+00
Zinc finger protein 292	ZNF292	O60281	2	2	0,00023305	4,63E+00
14-3-3 protein epsilon	YWHAE	P62258	2	10	4,7366E-36	4,61E+00
60S ribosomal protein L24	RPL24	P83731	3	3	5,3556E-07	4,60E+00
Protein S100-A16	S100A16	Q96FQ6	1	2	2,1974E-06	4,58E+00
Keratin-81-like protein KRT121P	KRT121P	A6NCN2	1	18	3,855E-182	4,56E+00
Suprabasin	SBSN	E9PBV3	1	3	3,9812E-65	4,54E+00
Selenium-binding protein 1	SELENBP1	Q13228	9	9	7,5262E-28	4,53E+00
Acetyl-CoA carboxylase 1;Biotin carboxylase	ACACA	Q13085	3	131	0	4,52E+00
Programmed cell death 6-interacting protein	PDCD6IP	Q8WUM4	3	7	2,487E-22	4,51E+00
Pituitary homeobox 3	PITX3	O75364	1	2	2,5684E-25	4,50E+00
Beta-synuclein	SNCB	Q16143	1	2	1,7467E-06	4,49E+00
RNA-binding protein 48	RBM48	Q5RL73	1	2	9,7514E-12	4,48E+00
Cyclin-dependent kinase 8;Cyclin-dependent kinase 19	CDK8;CDK19	P49336	7	2	8,7019E-05	4,46E+00
WD repeat-containing protein 36	WDR36	Q8NI36	3	6	3,942E-13	4,45E+00
Sentrin-specific protease 6	SENP6	Q9GZR1	10	5	1,4221E-16	4,44E+00
Histone RNA hairpin-binding protein	SLBP	Q14493	6	6	4,5535E-15	4,42E+00
Adapter molecule crk	CRK	P46108	2	5	2,606E-18	4,42E+00
Protein regulator of cytokinesis 1	PRC1	O43663	8	3	2,3894E-09	4,39E+00

Dermcidin;Survival-promoting peptide;DCD-1	DCD	P81605	2	5	1,2596E-45	4,39E+00
Regulator of chromosome condensation	RCC1	P18754	8	4	2,8059E-14	4,39E+00
Nuclear factor 1 C-type	NFIC	P08651	6	4	1,2747E-60	4,39E+00
40S ribosomal protein S18	RPS18	P62269	2	3	2,6686E-06	4,39E+00
Homeobox and leucine zipper protein Homez	HOMEZ	Q8IX15	2	2	7,6686E-05	4,39E+00
CWF19-like protein 2	CWF19L2	Q2TBE0	5	11	9,2864E-94	4,34E+00
Insulin gene enhancer protein ISL-2;Insulin gene enhancer protein ISL-1	ISL2;ISL1	Q96A47	6	2	2,7172E-07	4,33E+00
RNA-binding protein 4;RNA-binding protein 4B	RBM4;RBM4B	Q9BWF3	10	4	6,2989E-11	4,31E+00
40S ribosomal protein S25	RPS25	P62851	1	4	4,4193E-09	4,31E+00
Centrosomal protein POC5	POC5	Q8NA72	5	2	3,0584E-11	4,29E+00
Peptidyl-prolyl cis-trans isomerase E;Peptidyl-prolyl cis-trans isomerase	PPIE	Q9UNP9	6	3	6,1837E-11	4,28E+00
Gamma-glutamylcyclotransferase	GGCT	O75223	6	3	1,8829E-12	4,26E+00
Immunoglobulin J chain	IGJ	P01591	4	2	8,8149E-11	4,25E+00
60S ribosomal protein L6	RPL6	Q02878	4	2	1,9015E-38	4,22E+00
Prolactin-inducible protein	PIP	P12273	1	4	8,2827E-13	4,21E+00
Thioredoxin-like protein 1	TXNL1	O43396	2	12	1,3654E-89	4,20E+00
Zinc finger protein 629	ZNF629	Q9UEG4	26	8	7,9563E-19	4,18E+00
Glutamate dehydrogenase 1, mitochondrial;Glutamate dehydrogenase 2, mitochondrial;Glutamate dehydrogenase	GLUD1;GLUD2	P00367	7	2	4,9917E-06	4,15E+00
Protein S100-A14	S100A14	Q9HCY8	1	2	1,0057E-10	4,14E+00
Protein SDE2 homolog	SDE2	Q6IQ49	4	4	5,4987E-22	4,12E+00
Nitric oxide synthase-interacting protein	NOSIP	Q9Y314	1	5	1,7747E-19	4,11E+00
GON-4-like protein	GON4L	Q3T8J9	5	5	2,0681E-15	4,09E+00
RAD51-associated protein 1	RAD51AP1	Q96B01	6	3	7,0739E-10	4,08E+00
Exocyst complex component 4	EXOC4	Q96A65	6	11	1,2734E-61	4,08E+00
Histone-lysine N-methyltransferase SETD7	SETD7	Q8WTS6	3	7	6,1605E-54	4,08E+00
mRNA export factor	RAE1	P78406	5	6	4,9167E-28	4,06E+00
General transcription factor IIF subunit 2	GTF2F2	P13984	1	2	0,0015434	4,06E+00
Inactive phospholipase C-like protein 1	PLCL1	Q15111	2	2	0,0024801	4,05E+00
Histone-lysine N-methyltransferase EHMT2	EHMT2	Q96KQ7	12	2	0,0015779	4,05E+00
Nuclear inhibitor of protein phosphatase 1;Activator of RNA decay	PPP1R8	Q12972	4	3	2,0717E-14	4,04E+00
Fanconi anemia group J protein	BRIP1	Q9BX63	3	10	5,7289E-40	4,04E+00

Eukaryotic translation initiation factor 2 subunit 1	EIF2S1	P05198	3	3	1,1199E-16	4,04E+00
Farnesyl pyrophosphate synthase	FDPS	P14324	2	2	1,0066E-05	4,03E+00
DNA repair protein RAD50	RAD50	Q92878	5	4	3,4477E-09	4,02E+00
Breakpoint cluster region protein	BCR	P11274	14	7	2,0515E-14	4,02E+00
Chromodomain-helicase-DNA-binding protein 7	CHD7	Q9P2D1	9	11	1,2707E-20	4,02E+00
Regulator of nonsense transcripts 3B	UPF3B	Q9BZI7	3	2	0,00039364	4,02E+00
Glyceraldehyde-3-phosphate dehydrogenase, testis-specific	GAPDHS	O14556	1	2	5,3312E-05	4,02E+00
Neurofilament light polypeptide	NEFL	P07196	1	4	1,9742E-11	4,00E+00
Apoptosis inhibitor 5	API5	Q9BZZ5	9	17	1,068E-162	4,00E+00

IX. REFERENCES

Abedin, M. J., Imai, N., Rosenberg, M. E. and Gupta, S. (2011). Identification and characterization of sall1-expressing cells present in the adult mouse kidney. *Nephron - Exp. Nephrol.* **119**, 75–82.

Badenhorst, P., Finch, J. T. and Travers, A. a (2002). Tramtrack co-operates to prevent inappropriate neural development in Drosophila. *Mech. Dev.* **117**, 87–101.

Barembaum, M. and Bronner-Fraser, M. (2004). A novel spalt gene expressed in branchial arches affects the ability of cranial neural crest cells to populate sensory ganglia. *Neuron Glia Biol* **1**, 57–63.

Barrio, R. and de Celis, J. F. (2004). Regulation of spalt expression in the Drosophila wing blade in response to the Decapentaplegic signaling pathway. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6021–6026.

Barrio, R., Shea, M. J., Carulli, J., Lipkow, K., Gaul, U., Frommer, G., Schuh, R., Jäckle, H. and Kafatos, F. C. (1996). The spalt-related gene of Drosophila melanogaster is a member of an ancient gene family, defined by the adjacent, region-specific homeotic gene spalt. *Dev. Genes Evol.* **206**, 315–25.

Barrio, R., de Celis, J. F., Bolshakov, S. and Kafatos, F. C. (1999). Identification of regulatory regions driving the expression of the Drosophila spalt complex at different developmental stages. *Dev. Biol.* **215**, 33–47.

Bate, M. and Arias, A. M. (1991). The embryonic origin of imaginal discs in Drosophila. *Development* **112**, 755–761.

Becker, J., Barysch, S. V, Karaca, S., Dittner, C., Hsiao, H.-H., Diaz, M. B., Herzig, S., Urlaub, H. and Melchior, F. (2013). Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* **20**, 525–531.

Berdnik, D., Favaloro, V. and Liqun, L. (2012). The SUMO Protease Verloren Regulates Dendrite and Axon Targeting in Olfactory Projection Neurons. *J. Neurosci.* **32**, 8331–8340.

Bernier-Villamor, V., Sampson, D. a, Matunis, M. J. and Lima, C. D. (2002). Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* **108**, 345–56.

Bhaskar, V., Valentine, S. a and Courey, a J. (2000). A functional interaction between dorsal and components of the Smt3 conjugation machinery. *J Biol Chem* **275**, 4033–4040.

Bhaskar, V., Smith, M. and Courey, A. J. (2002). Conjugation of Smt3 to dorsal may potentiate the Drosophila immune response. *Mol Cell Biol* **22**, 492–504.

Borozdin, W., Boehm, D., Leipoldt, M., Wilhelm, C., Reardon, W., Clayton-Smith, J., Becker, K., Muhlendyck, H., Winter, R., Giray, O., et al. (2004). SALL4 deletions are a common cause of Okihiro and acro-renal-ocular syndromes and confirm haploinsufficiency as the pathogenic mechanism. *J. Med. Genet.* **41**, e113.

Borozdin, W., Steinmann, K., Albrecht, B., Bottani, A., Devriendt, K., Leipoldt, M. and Kohlhase, J. (2006). Detection of heterozygous SALL1 deletions by quantitative real time PCR proves the contribution of a SALL1 dosage effect in the pathogenesis of Townes-Brocks syndrome. *Hum. Mutat.* **27**, 211–2.

Borozdin, W., Graham Jr., J. M., Bohm, D., Bamshad, M. J., Spranger, S., Burke, L., Leipoldt, M. and Kohlhase, J. (2007). Multigene deletions on chromosome 20q13.13-q13.2 including SALL4 result in an expanded phenotype of Okihiro syndrome plus developmental delay. *Hum. Mutat.* **28**, 830.

Botzenhart, E. M., Green, A., Ilyina, H., König, R., Lowry, R. B., Lo, I. F. M., Shohat, M., Burke, L., McGaughran, J., Chafai, R., et al. (2005). SALL1 mutation analysis in Townes-Brocks syndrome: twelve novel mutations and expansion of the phenotype. *Hum. Mutat.* **26**, 282.

Botzenhart, E. M., Bartalini, G., Blair, E., Brady, A. F., Elmslie, F., Chong, K. L., Christy, K., Torres-Martinez, W., Danesino, C., Deardorff, M. A., et al. (2007). Townes-Brocks syndrome: twenty novel SALL1 mutations in sporadic and familial cases and refinement of the SALL1 hot spot region. *Hum. Mutat.* **28**, 204–5.

Brand, a H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–15.

Breuer, K., Foroushani, a. K., Laird, M. R., Chen, C., Sribnaia, a., Lo, R., Winsor, G. L., Hancock, R. E. W., Brinkman, F. S. L. and Lynn, D. J. (2013). InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Res.* **41**, D1228–D1233.

Bruderer, R., Tatham, M. H., Plechanovova, A., Matic, I., Garg, A. K. and Hay, R. T. (2011). Purification and identification of endogenous polySUMO conjugates. *EMBO Rep.* **12**, 142–148.

Buaas, F. W., Gardiner, J. R., Clayton, S., Val, P. and Swain, A. (2012). In vivo evidence for the crucial role of SF1 in steroid-producing cells of the testis, ovary and adrenal gland. *Development* **139**, 4561–70.

Buck, A., Archangelo, L., Dixkens, C. and Kohlhase, J. (2000). Molecular cloning, chromosomal localization, and expression of the murine SALL1 ortholog Sall1. *Cytogenet. Cell Genet.* **89**, 150–153.

Buck, A., Kispert, A. and Kohlhase, J. (2001). Embryonic expression of the murine homologue of SALL1, the gene mutated in Townes--Brocks syndrome. *Mech. Dev.* **104**, 143–6.

Bylebyl, G. R., Belichenko, I. and Johnson, E. S. (2003). The SUMO Isopeptidase Ulp2 Prevents Accumulation of SUMO Chains in Yeast. *J. Biol. Chem.* **278**, 44113–44120.

Camp, E., Hope, R., Kortschak, R. D., Cox, T. C. and Lardelli, M. (2003). Expression of three spalt (sal) gene homologues in zebrafish embryos. *Dev. Genes Evol.* **213**, 35–43.

Cantera, R., Luer, K., Rusten, T. E., Barrio, R., Kafatos, F. C. and Technau, G. M. (2002). Mutations in spalt cause a severe but reversible neurodegenerative phenotype in the embryonic central nervous system of Drosophila melanogaster. *Development* **129**, 5577–5586.

Cappadocia, L., Pichler, A. and Lima, C. D. (2015). Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase. *Nat. Struct. Mol. Biol.*

Casanova, J. (1989). Mutations in the spalt gene of Drosophila cause ectopic expression of Ultrabithorax and Sex combs reduced. *Dev. Biol.* **198**, 137–140.

Chen, W.-Y., Lee, W.-C., Hsu, N.-C., Huang, F. and Chung, B.-C. (2004). SUMO modification of repression domains modulates function of nuclear receptor 5A1 (steroidogenic factor-1). *J. Biol. Chem.* **279**, 38730–5.

Christianson, a M., King, D. L., Hatzivassiliou, E., Casas, J. E., Hallenbeck, P. L., Nikodem, V. M., Mitsialis, S. a and Kafatos, F. C. (1992). DNA binding and heteromerization of the Drosophila transcription factor chorion factor 1/ultraspiracle. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11503–7.

Da Silva-Ferrada, E., Xolalpa, W., Lang, V., Aillet, F., Martin-Ruiz, I., de la Cruz-Herrera, C. F., Lopitz-Otsoa, F., Carracedo, A., Goldenberg, S. J., Rivas, C., et al. (2013). Analysis of SUMOylated proteins using SUMO-traps. *Sci. Rep.* **3**, 2–7.

Daniel, J. and Liebau, E. (2014). The ufm1 cascade. *Cells* **3**, 627–38.

De Celis, J. F. (2003). Pattern formation in the Drosophila wing: The development of the veins. *BioEssays* **25**, 443–451.

De Celis, J. F. and Barrio, R. (2000). Function of the spalt/spalt-related gene complex in positioning the veins in the Drosophila wing. *Mech. Dev.* **91**, 31–41.

de Celis, J. F. and Barrio, R. (2009). Regulation and function of Spalt proteins during animal development. *Int. J. Dev. Biol.* **53**, 1385–1398.

de Navas, L. F., Reed, H., Akam, M., Barrio, R., Alonso, C. R. and Sanchez-Herrero, E. (2011). Integration of RNA processing and expression level control modulates the function of the Drosophila Hox gene Ultrabithorax during adult development. *Development* **138**, 107–116.

Denison, C. (2005). A Proteomic Strategy for Gaining Insights into Protein Sumoylation in Yeast. *Mol. Cell. Proteomics* **4**, 246–254.

Dingar, D., Kalkat, M., Chan, P.-K., Srikumar, T., Bailey, S. D., Tu, W. B., Coyaud, E., Ponzielli, R., Kolyar, M., Jurisica, I., et al. (2014). BioID identifies novel c-MYC interacting partners in cultured cells and xenograft tumors. *J. Proteomics* 1–17.

Dubrovsky, E. B. (2005). Hormonal cross talk in insect development. *Trends Endocrinol. Metab.* **16**, 6–11.

Dundr, M. and Misteli, T. (2001). Functional architecture in the cell nucleus. *Biochem. J.* **356**, 297–310.

Eifler, K. and Vertegaal, A. C. O. (2015). Mapping the SUMOylated landscape. *FEBS J.* n/a–n/a.

Eisenhardt, N., Chaugule, V. K., Koidl, S., Droescher, M., Dogan, E., Rettich, J., Sutinen, P., Imanishi, S. Y., Hofmann, K., Palvimo, J. J., et al. (2015). A new vertebrate SUMO enzyme family reveals insights into SUMO-chain assembly. *Nat. Struct. Mol. Biol.*

Epps, J. L. and Tanda, S. (1998). The Drosophila semushi mutation blocks nuclear import of Bicoid during embryogenesis. *Curr. Biol.* **8**, 1277–S2.

Farrell, E. R., Tosh, G., Church, E. and Münsterberg, a E. (2001). Cloning and expression of CSAL2, a new member of the spalt gene family in chick. *Mech. Dev.* **102**, 227–30.

Flotho, A. and Melchior, F. (2013). Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* **82**, 357–85.

Franco, M., Seyfried, N. T., Brand, A. H., Peng, J. and Mayor, U. (2011). A novel strategy to isolate ubiquitin conjugates reveals wide role for ubiquitination during neural development. *Mol. Cell. Proteomics* **10**, M110.002188.

Ganesan, A. K., Kho, Y., Sung, C. K., Chen, Y., Zhao, Y. and White, M. a. (2007). Broad spectrum identification of SUMO substrates in melanoma cells. *Proteomics* **7**, 2216–2221.

Gates, J., Nowotarski, S. H., Yin, H., Mahaffey, J. P., Bridges, T., Herrera, C., Homem, C. C. F., Janody, F., Montell, D. J. and Peifer, M. (2009). Enabled and Capping protein play important roles in shaping cell behavior during Drosophila oogenesis. *Dev. Biol.* **333**, 90–107.

Gonzalez, I., Mateos-Langerak, J., Thomas, A., Cheutin, T. and Cavalli, G. (2014). Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. *Mol. Cell* **54**, 485–499.

González, M., Martín-Ruíz, I., Jiménez, S., Pirone, L., Barrio, R. and Sutherland, J. D. (2011). Generation of stable Drosophila cell lines using multicistronic vectors. *Sci. Rep.* **1**,.

Gullberg, M., Gústafsdóttir, S. M., Schallmeiner, E., Jarvius, J., Bjarnegård, M., Betsholtz, C., Landegren, U. and Fredriksson, S. (2004). Cytokine detection by antibody-based proximity ligation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8420–8424.

Guo, D., Li, M., Zhang, Y., Yang, P., Eckenrode, S., Hopkins, D., Zheng, W., Purohit, S., Podolsky, R. H., Muir, A., et al. (2004). A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. *Nat. Genet.* **36**, 837–41.

Hannich, J. T., Lewis, A., Kroetz, M. B., Li, S.-J. J., Heide, H., Emili, A. and Hochstrasser, M. (2005). Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. *J.Biol.Chem.* **280**, 4102–4110.

Hari, K. L., Hari, K. L., Cook, K. R., Cook, K. R., Karpen, G. H. and Karpen, G. H. (2001). Drosophila Su(var)2-10. *Genes Dev.* 1334–1348.

Hay, B. a, Maile, R. and Rubin, G. M. (1997). P element insertion-dependent gene activation in the Drosophila eye. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5195–5200.

Hecker, C.-M., Rabiller, M., Haglund, K., Bayer, P. and Dikic, I. (2006). Specification of SUMO1- and SUMO2-interacting Motifs. *J. Biol. Chem.* **281**, 16117–16127.

Hendriks, I. a., D'Souza, R. C., Chang, J.-G., Mann, M. and Vertegaal, A. C. O. (2015). System-wide identification of wild-type SUMO-2 conjugation sites. *Nat. Commun.* **6**, 7289.

Hietakangas, V., Anckar, J., Blomster, H. a, Fujimoto, M., Palvimo, J. J., Nakai, A. and Sistonen, L. (2006). PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 45–50.

Ismail, I. H., Gagné, J. P., Caron, M. C., McDonald, D., Xu, Z., Masson, J. Y., Poirier, G. G. and Hendzel, M. J. (2012). CBX4-mediated SUMO modification regulates BMI1 recruitment at sites of DNA damage. *Nucleic Acids Res.* **40**, 5497–5510.

Kagey, M. H., Melhuish, T. a. and Wotton, D. (2003). The polycomb protein Pc2 is a SUMO E3. *Cell* **113**, 127–137.

Kagey, M. H., Melhuish, T. a, Powers, S. E. and Wotton, D. (2005). Multiple activities contribute to Pc2 E3 function. *EMBO J.* **24**, 108–119.

Kanakousaki, K. and Gibson, M. C. (2012). A differential requirement for SUMOylation in proliferating and non-proliferating cells during Drosophila development. *Development* **139**, 2751–2762.

Kelberman, D., Islam, L., Lakowski, J., Bacchelli, C., Chanudet, E., Lescai, F., Patel, A., Stupka, E., Buck, A., Wolf, S., et al. (2014). Mutation of SALL2 causes recessive ocular coloboma in humans and mice. *Hum. Mol. Genet.* **23**, 2511–2526.

Kiefer, S. M., McDill, B. W., Yang, J. and Rauchman, M. (2002). Murine Sall1 represses transcription by recruiting a histone deacetylase complex. *J. Biol. Chem.* **277**, 14869–14876.

Knipscheer, P., van Dijk, W. J., Olsen, J. V, Mann, M. and Sixma, T. K. (2007). Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation. *EMBO J.* **26**, 2797–2807.

Knipscheer, P., Flotho, A., Klug, H., Olsen, J. V., van Dijk, W. J., Fish, A., Johnson, E. S., Mann, M., Sixma, T. K. and Pichler, A. (2008). Ubc9 Sumoylation Regulates SUMO Target Discrimination. *Mol. Cell* **31**, 371–382.

Kohlhase, J., Wischermann, A., Reichenbach, H., Froster, U. and Engel, W. (1998). Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome.

Kohlhase, J., Heinrich, M., Schubert, L., Liebers, M., Kispert, A., Laccone, F., Turnpenny, P., Winter, R. M. and Reardon, W. (2002). Okihiro syndrome is caused by SALL4 mutations. *Hum. Mol. Genet.* **11**, 2979–2987.

Komatsu, M., Chiba, T., Tatsumi, K., Iemura, S., Tanida, I., Okazaki, N., Ueno, T., Kominami, E., Natsume, T. and Tanaka, K. (2004). A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO J.* **23**, 1977–1986.

Kühnlein, R. P. and Schuh, R. (1996). Dual function of the region-specific homeotic gene spalt during Drosophila tracheal system development. *Development* **122**, 2215–23.

Kühnlein, R. P., Frommer, G., Friedrich, M., Gonzalez-Gaitan, M., Weber, A., Wagner-Bernholz, J. F., Gehring, W. J., Jäckle, H. and Schuh, R. (1994). spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the Drosophila embryo. *EMBO J.* **13**, 168–79.

Lachance, P. E. D., Miron, M., Raught, B., Lasko, P. and Sonenberg, N. (2002). Phosphorylation of Eukaryotic Translation Initiation Factor 4E Is Critical for Growth Phosphorylation of Eukaryotic Translation Initiation Factor 4E Is Critical for Growth. *Mol. Cell. Biol.* **22**, 1656–1663.

Landecker, H. L., Sinclair, D. a R. and Brock, H. W. (1994). Screen for enhancers of Polycomb and Polycomblike in Drosophila melanogaster. *Dev. Genet.* **15**, 425–434.

Lauberth, S. M. and Rauchman, M. (2006). A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J. Biol. Chem.* **281**, 23922–23931.

Lauberth, S. M., Bilyeu, A. C., Firulli, B. A., Kroll, K. L. and Rauchman, M. (2007). A phosphomimetic mutation in the Sall1 repression motif disrupts recruitment of the nucleosome remodeling and deacetylase complex and repression of Gbx2. *J Biol Chem* **282**, 34858–34868.

Lehembre, F., Badenhorst, P., Müller, S., Travers, a, Schweisguth, F. and Dejean, a (2000). Covalent modification of the transcriptional repressor tramtrack by the ubiquitin-related protein Smt3 in Drosophila flies. *Mol. Cell. Biol.* **20**, 1072–1082.

Lemaire, K., Moura, R. F., Granvik, M., Igoillo-Esteve, M., Hohmeier, H. E., Hendrickx, N., Newgard, C. B., Waelkens, E., Cnop, M. and Schuit, F. (2011). Ubiquitin fold modifier 1 (UFM1) and its target UFBP1 protect pancreatic beta cells from ER stress-induced apoptosis. *PLoS One* **6**, e18517.

Li, C., Guo, M., Borczuk, A., Powell, C. A., Wei, M., Thaker, H. M., Friedman, R., Klein, U. and Tycko, B. (2002). Gene Expression in Wilms ' Tumor Mimics the Earliest Committed Stage in the Metanephric Mesenchymal-Epithelial Transition. **160**, 2181–2190.

Li, B., Zhou, J., Liu, P., Hu, J., Jin, H., Shimono, Y., Takahashi, M. and Xu, G. (2007). Polycomb protein Cbx4 promotes SUMO modification of *de novo* DNA methyltransferase Dnmt3a. *Biochem. J.* **405**, 369–378.

Liu, H., Adler, A. S., Segal, E. and Chang, H. Y. (2007). A transcriptional program mediating entry into cellular quiescence. *PLoS Genet.* **3**, 0996–1008.

Long, X. and Griffith, L. C. (2000). Identification and Characterization of a SUMO-1 Conjugation System That Modifies Neuronal Calcium/Calmodulin-dependent Protein Kinase II in Drosophila melanogaster. *J. Biol. Chem.* **275**, 40765–40776.

Ma, Y., Li, D., Chai, L., Luciani, A. M., Ford, D., Morgan, J. and Maizel, A. L. (2001). *Cloning and Characterization of Two Promoters for the Human HSAL2 Gene and Their Transcriptional Repression by the Wilms Tumor Suppressor Gene Product*.

MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M. and Sadowski, P. D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. *Mol. Cell. Biol.* **29**, 714–725.

Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. (1997). A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97–107.

Matic, I., van Hagen, M., Schimmel, J., Macek, B., Ogg, S. C., Tatham, M. H., Hay, R. T., Lamond, A. I., Mann, M. and Vertegaal, A. C. O. (2008). In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol. Cell. Proteomics* **7**, 132–144.

Matic, I., Schimmel, J., Hendriks, I. a., van Santen, M. a., van de Rijke, F., van Dam, H., Gnad, F., Mann, M. and Vertegaal, A. C. O. (2010). Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif. *Mol. Cell* **39**, 641–652.

Merrill, J. C., Melhuish, T. a., Kagey, M. H., Yang, S. H., Sharrocks, A. D. and Wotton, D. (2010). A role for non-covalent SUMO interaction motifs in Pc2/CBX4 E3 activity. *PLoS One* **5**,.

Miller, E. M., Hopkin, R., Bao, L. and Ware, S. M. (2011). Implications for Genotype – Phenotype Predictions in Townes – Brocks Syndrome : Case Report of a Novel SALL1 Deletion and Review of the Literature. *Am. J. Med. Genet.* 1–8.

Minty, a., Dumont, X., Kaghad, M. and Caput, D. (2000). Covalent Modification of p73 by SUMO-1: TWO-HYBRID SCREENING WITH p73 IDENTIFIES NOVEL SUMO-1-INTERACTING PROTEINS AND A SUMO-1 INTERACTION MOTIF. *J. Biol. Chem.* **275**, 36316–36323.

Mirth, C., Truman, J. W. and Riddiford, L. M. (2005). The role of the prothoracic gland in determining critical weight for metamorphosis in Drosophila melanogaster. *Curr. Biol.* **15**, 1796–1807.

Monribot-Villanueva, J., Juárez-Uribe, R. A., Palomera-Sánchez, Z., Gutiérrez-Aguiar, L., Zurita, M., Kennison, J. a. and Vázquez, M. (2013). TnaA, an SP-RING Protein, Interacts with Osa, a Subunit of the Chromatin Remodeling Complex BRAHMA and with the SUMOylation Pathway in Drosophila melanogaster. *PLoS One* **8**,.

Nagai, S., Davoodi, N. and Gasser, S. M. (2011). Nuclear organization in genome stability: SUMO connections. *Cell Res.* **21**, 474–485.

Nayak, A. and Müller, S. (2014). SUMO-specific proteases/isopeptidases: SENPs and beyond. *Genome Biol.* **15**, 422.

Netzer, C., Rieger, L., Brero, a, Zhang, C. D., Hinzke, M., Kohlhase, J. and Bohlander, S. K. (2001). SALL1, the gene mutated in Townes-Brocks syndrome, encodes a transcriptional repressor which interacts with TRF1/PIN2 and localizes to pericentromeric heterochromatin. *Hum. Mol. Genet.* **10**, 3017–3024.

Netzer, C., Bohlander, S. K., Rieger, L., Müller, S. and Kohlhase, J. (2002). Interaction of the developmental regulator SALL1 with UBE2I and SUMO-1. *Biochem. Biophys. Res.*

Commun. **296**, 870–876.

Netzer, C., Bohlander, S. K., Hinzke, M., Chen, Y. and Kohlhase, J. (2006). Defining the heterochromatin localization and repression domains of SALL1. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1762**, 386–391.

Nie, M., Xie, Y., Loo, J. a. and Courey, A. J. (2009). Genetic and proteomic evidence for roles of Drosophila SUMO in cell cycle control, Ras signaling, and early pattern formation. *PLoS One* **4**,.

Nishinakamura, R., Matsumoto, Y., Nakao, K., Nakamura, K., Sato, A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Scully, S., Lacey, D. L., et al. (2001). Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development. *Development* **128**, 3105–3115.

Ohsako, S. and Takamatsu, Y. (1999). Identification and characterization of a Drosophila Homologue of the Yeast UBC9 and hus5 Genes. *J. Biochem.* **125**, 230–235.

Organista, M. F., Martín, M., de Celis, J. M., Barrio, R., López-Varea, A., Esteban, N., Casado, M. and de Celis, J. F. (2015). The Spalt Transcription Factors Generate the Transcriptional Landscape of the Drosophila melanogaster Wing Pouch Central Region. *PLOS Genet.* **11**, e1005370.

Owerbach, D., McKay, E. M., Yeh, E. T. H., Gabbay, K. H. and Bohren, K. M. (2005). A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem. Biophys. Res. Commun.* **337**, 517–520.

Pichler, A., Gast, A., Seeler, J. S., Dejean, A. and Melchior, F. (2002). The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* **108**, 109–120.

Pirrotta, V. and Li, H. B. (2012). A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* **22**, 101–109.

Rodriguez, M. S., Dargemont, C. and Hay, R. T. (2001). SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting. *J. Biol. Chem.* **276**, 12654–12659.

Roux, K. J., Kim, D. I., Raida, M. and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810.

Rusten, T. E., Cantera, R., Urban, J., Technau, G., Kafatos, F. C. and Barrio, R. (2001). Spalt modifies EGFR-mediated induction of chordotonal precursors in the embryonic PNS of Drosophila promoting the development of oenocytes. *Development* **128**, 711–22.

Sánchez, J., Talamillo, A., Lopitz-Otsoa, F., Pérez, C., Hjerpe, R., Sutherland, J. D., Herboso, L., Rodríguez, M. S. and Barrio, R. (2010). Sumoylation modulates the activity of spalt-like proteins during wing development in Drosophila. *J. Biol. Chem.* **285**, 25841–25849.

Sánchez, J., Talamillo, A., González, M., Sánchez-Pulido, L., Jiménez, S., Pirone, L., Sutherland, J. D. and Barrio, R. (2011). Drosophila Sal and Salr are transcriptional repressors. *Biochem. J.* **438**, 437–445.

Schimmel, J., Eifler, K., Sigursson, J., Cuijpers, S. a G., Hendriks, I. a., Verlaan-de Vries, M., Kelstrup, C. D., Francavilla, C., Medema, R. H., Olsen, J. V., et al. (2014). Uncovering SUMOylation dynamics during cell-cycle progression reveals foxM1 as a key mitotic SUMO target protein. *Mol. Cell* **53**, 1053–1066.

Schulz, S., Chachami, G., Kozaczkiewicz, L., Winter, U., Stankovic-Valentin, N., Haas, P., Hofmann, K., Urlaub, H., Ovaa, H., Wittbrodt, J., et al. (2012). Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. *EMBO Rep.* **13**, 930–938.

Shin, E. J., Shin, H. M., Nam, E., Kim, W. S., Kim, J.-H., Oh, B.-H. and Yun, Y. (2012). DeSUMOylating isopeptidase: a second class of SUMO protease. *EMBO Rep.* **13**, 339–46.

Smith, M., Bhaskar, V., Fernandez, J. and Courey, A. J. (2004). *Drosophila* Ulp1, a Nuclear Pore-associated SUMO Protease, Prevents Accumulation of Cytoplasmic SUMO Conjugates. *J. Biol. Chem.* **279**, 43805–43814.

Smith, M., Turki-Judeh, W. and Courey, A. J. (2012). SUMOylation in Drosophila Development. *Biomolecules* **2**, 331–349.

Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G., et al. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000.

Song, J. (2005). Small Ubiquitin-like Modifier (SUMO) Recognition of a SUMO Binding Motif: A REVERSAL OF THE BOUND ORIENTATION. *J. Biol. Chem.* **280**, 40122–40129.

Sorokin, a V, Kim, E. R. and Ovchinnikov, L. P. (2009). Proteasome system of protein degradation and processing. *Biochem. Biokhimiia* **74**, 1411–42.

Sung, C. K., Li, D., Andrews, E., Drapkin, R. and Benjamin, T. (2013). Promoter methylation of the SALL2 tumor suppressor gene in ovarian cancers. *Mol. Oncol.* **7**, 419–427.

Sweetman, D. and Münsterberg, A. (2006). The vertebrate spalt genes in development and disease. *Dev. Biol.* **293**, 285–293.

Sweetman, D., Smith, T., Farrell, E. R., Chantry, A. and Munsterberg, A. (2003). The Conserved Glutamine-rich Region of Chick Csal1 and Csal3 Mediates Protein Interactions with Other Spalt Family Members: IMPLICATIONS FOR TOWNES-BROCKS SYNDROME. *J. Biol. Chem.* **278**, 6560–6566.

Takanaka, Y. and Courey, A. J. (2005). SUMO enhances Vestigial function during wing morphogenesis. *Mech. Dev.* **122**, 1130–1137.

Talamillo, A., Sánchez, J. and Barrio, R. (2008a). Functional analysis of the SUMOylation pathway in Drosophila. *Biochem. Soc. Trans.* **36**, 868–73.

Talamillo, A., Sánchez, J., Cantera, R., Pérez, C., Martín, D., Caminero, E. and Barrio, R. (2008b). Smt3 is required for Drosophila melanogaster metamorphosis. *Development* **135**, 1659–1668.

Talamillo, A., Herboso, L., Pirone, L., Pérez, C., González, M., Sánchez, J., Mayor, U., Lopitz-Otsoa, F., Rodriguez, M. S., Sutherland, J. D., et al. (2013). Scavenger Receptors Mediate the Role of SUMO and Ftz-f1 in Drosophila Steroidogenesis. *PLoS Genet.* **9**,.

Tammsalu, T., Matic, I., Jaffray, E. G., Ibrahim, A. F. M., Tatham, M. H. and Hay, R. T. (2014). Proteome-wide identification of SUMO2 modification sites. *Sci. Signal.* **7**, rs2.

Tatham, M. H., Jaffray, E., Vaughan, O. a., Desterro, J. M. P., Botting, C. H., Naismith, J. H. and Hay, R. T. (2001). Polymeric Chains of SUMO-2 and SUMO-3 are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9. *J. Biol. Chem.* **276**, 35368–35374.

Tatham, M. H., Geoffroy, M.-C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E. G., Palvimo, J. J. and Hay, R. T. (2008). RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* **10**, 538–546.

Tatsumi, K., Yamamoto-Mukai, H., Shimizu, R., Waguri, S., Sou, Y.-S., Sakamoto, A., Taya, C., Shitara, H., Hara, T., Chung, C. H., et al. (2011). The Ufm1-activating enzyme Uba5 is indispensable for erythroid differentiation in mice. *Nat. Commun.* **2**, 181.

Tirard, M., Hsiao, H., Nikolov, M., Urlaub, H., Melchior, F. and Brose, N. (2012). In vivo localization and identi fi cation of SUMOylated proteins in the brain of His 6 -HA-SUMO1 knock-in mice. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21122–21127.

Toker, A. S., Teng, Y., Ferreira, H. B., Emmons, S. W. and Chalfie, M. (2003). The Caenorhabditis elegans spalt-like gene sem-4 restricts touch cell fate by repressing the selector Hox gene egl-5 and the effector gene mec-3. *Development* **130**, 3831–3840.

Ureña, E., Pirone, L., Chafino, S., Pérez, C., Sutherland, J. D., Lang, V., Rodriguez, M. S., Lopitz-Otsoa, F., Blanco, F. J., Barrio, R., et al. (2015). Evolution of SUMO Function and Chain Formation in Insects. *Mol. Biol. Evol.* msv242.

van der Veen, A. G. and Ploegh, H. L. (2012). Ubiquitin-Like Proteins. *Annu. Rev. Biochem.* **81**, 323–357.

Vertegaal, A. C. O., Ogg, S. C., Jaffray, E., Rodriguez, M. S., Hay, R. T., Andersen, J. S., Mann, M. and Lamond, A. I. (2004). A proteomic study of SUMO-2 target proteins. *J. Biol. Chem.* **279**, 33791–33798.

Wang, J., Wang, S. and Li, S. (2014). Sumoylation modulates 20-hydroxyecdysone signaling by maintaining USP protein levels in Drosophila. *Insect Biochem. Mol. Biol.* **54**, 80–88.

Weatherbee, S. D., Halder, G., Kim, J., Hudson, A. and Carroll, S. (1998). Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere. *Genes Dev.* **12**, 1474–82.

Wei, W., Yang, P., Pang, J., Zhang, S., Wang, Y., Wang, M. H., Dong, Z., She, J. X. and Wang, C. Y. (2008). A stress-dependent SUMO4 sumoylation of its substrate proteins. *Biochem. Biophys. Res. Commun.* **375**, 454–459.

Werner, A., Flotho, A. and Melchior, F. (2012). The RanBP2/RanGAP1∗SUMO1/Ubc9 Complex Is a Multisubunit SUMO E3 Ligase. *Mol. Cell* **46**, 287–298.

White, R. a and Wilcox, M. (1984). Protein products of the bithorax complex in Drosophila. *Cell* **39**, 163–171.

Wohlschlegel, J. A., Johnson, E. S., Reed, S. I. and Yates 3rd, J. R. (2004). Global analysis of protein sumoylation in Saccharomyces cerevisiae. *J Biol Chem* **279**, 45662–45668.

Wu, Q., Chen, X., Zhang, J., Loh, Y. H., Low, T. Y., Zhang, W., Zhang, W., Sze, S. K., Lim, B. and Ng, H. H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J. Biol. Chem.* **281**, 24090–24094.

Xi, P., Ding, D., Zhou, J., Wang, M. and Cong, Y.-S. (2013). DDRGK1 regulates NF-κB activity by modulating IκBα stability. *PLoS One* **8**, e64231.

Xiong, J. (2014). Sall4 SALL4 : Engine of Cell Stemness. *Curr. Gene Ther.* **14**, 400–411.

Xiong, J., Todorova, D., Su, N.-Y., Kim, J., Lee, P.-J., Shen, Z., Briggs, S. P. and Xu, Y. (2015). Stemness factor Sall4 is required for DNA damage response in embryonic stem cells. *J. Cell Biol.* **208**, 513–520.

Xu, X., Vatsyayan, J., Gao, C., Bakkenist, C. J. and Hu, J. (2010). HDAC2 promotes eIF4E sumoylation and activates mRNA translation gene specifically. *J. Biol. Chem.* **285**, 18139–18143.

Xu, Y., Plechanovová, A., Simpson, P., Marchant, J., Leidecker, O., Kraatz, S., Hay, R. T. and Matthews, S. J. (2014). Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. *Nat. Commun.* **5**,.

Yang, J., Gao, C., Chai, L. and Ma, Y. (2010). A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. *PLoS One* **5**, e10766.

Yang, X.-X., Sun, J.-Z., Li, F.-X., Wu, Y.-S., Du, H.-Y., Zhu, W., Li, X.-H. and Li, M. (2012a). Aberrant methylation and downregulation of sall3 in human hepatocellular carcinoma. *World J. Gastroenterol.* **18**, 2719–26.

Yang, F., Yao, Y., Jiang, Y., Lu, L., Ma, Y. and Dai, W. (2012b). Sumoylation is important for stability, subcellular localization, and transcriptional activity of SALL4, an essential stem cell transcription factor. *J. Biol. Chem.* **287**, 38600–38608.

Yoo, H. M., Kang, S. H., Kim, J. Y., Lee, J. E., Seong, M. W., Lee, S. W., Ka, S. H., Sou, Y.-S., Komatsu, M., Tanaka, K., et al. (2014). Modification of ASC1 by UFM1 Is Crucial for ERα Transactivation and Breast Cancer Development. *Mol. Cell* **56**, 261–74.

Yu, J., Zhu, T., Wang, Z., Zhang, H., Qian, Z., Xu, H., Gao, B. and Wang, W. (2007). A Novel Set of DNA Methylation Markers in Urine Sediments for Sensitive / Specific Detection of Bladder Cancer Human Cancer Biology A Novel Set of DNA Methylation Markers in Urine Sediments for Sensitive / Specific Detection of Bladder Cancer. **13**, 7296–7304.

Yunus, A. a. and Lima, C. D. (2009). Structure of the Siz/PIAS SUMO E3 Ligase Siz1 and Determinants Required for SUMO Modification of PCNA. *Mol. Cell* **35**, 669–682.

Zhang, Y.-Q. and Sarge, K. D. (2008). Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies. *J. Cell Biol.* **182**, 35–39.

Zhang, Y., Zhang, M., Wu, J., Lei, G. and Li, H. (2012). Transcriptional Regulation of the Ufm1 Conjugation System in Response to Disturbance of the Endoplasmic Reticulum Homeostasis and Inhibition of Vesicle Trafficking. *PLoS One* **7**, 1–11.

Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., Liu, Z., Zhao, Y., Xue, Y. and Ren, J. (2014). GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res.* **42**, W325–W330.

Zhong, S., Müller, S., Ronchetti, S., Freemont, P. S., Dejean, a and Pandolfi, P. P. (2000). Role of SUMO-1-modified PML in nuclear body formation. *Blood* **95**, 2748–2752.