UNIVERSITA' DEGLI STUDI DI NAPOLI "FEDERICO II"

RESEARCH DOCTORATE

in COMPUTATIONAL BIOLOGY and BIOINFORMATICS
XXV CYCLE

"Gene regulatory Network controlling anterior-posterior patterning of Central Nervous System and melanosomal logistics in Ciona intestinalis: a transcriptomic approach."

Candidate: Claudia Racioppi

Tutor   Coordinator
Dott.ssa Filomena Ristoratore   Prof. Sergio Cocozza

Co-Tutor
Dott. Diego di Bernardo

Academic Year 2011-2012

TABLE OF CONTENTS

CHAPTER 3 RESULTS

CHAPTER 4 CONCLUSIONS

APPENDIX

REFERENCES   **130**

ABSTRACT

Fibroblast growth factor (FGF) signaling plays an important role in neural induction and subsequent development throughout the vertebrates. Likewise, the FGF signaling pathway also functions in neural development in ascidians, that are considered as basal chordate sharing conserved molecular developmental mechanisms with vertebrates. FGF and Wnt signaling are involved in development of the pigment cells in the sensory organs of *C. intestinalis*. Nevertheless the FGF downstream targets, involved in pigment cell precursors formation, are not yet characterized and less is known about the patterning mechanisms in specification of the anteriormost nervous system in *Ciona*. Here, I employed a cell sorting, microarrays, and targeted manipulations to analyse the FGF signaling gene regulatory network controlling Central Nervous System (CNS) differentiation and pigmented cell formation in *C. intestinalis*. Microarray data analysis revealed that FGF signaling is required for setting the anterior-posterior CNS patterning in *Ciona*. My results indicate that when FGF signaling is blocked in Pigment Cell Precursors (PCPs), the most anterior cells change their fate, behaving like cells of the anterior CNS. Microarray data analysis produced hundreds genes differentially expressed when FGF signaling is pertubated in pigment cell lineage. A 'reverse engineered' approach has been employed to create a pigment cell lineage-specific transcriptional network from my microarray data set and neighbors analysis permitted to identify new potential candidates involved in FGF signaling regulatory architecture. Clustering algorithms permitted to group genes according to similar expression pattern; a bioinformatic tool was used to identify statistically the over-representation of conserved transcription factor binding sites on co-expressed gene promoter regions. This analysis allowed to computationally recover putative transcription factors involved in the FGF signaling regulatory network. Furthermore from microarray data analysis, I identified new specific markers of *Ciona* pigmented cell lineage and several genes that may function as key factors to prime the melanosomal logistic, including biogenesis and transport. Among these, I focused on *Ci-Rab38/32/rab-rp1/ltd*, the only orthologue of two highly conserved small Rab GTPase, RAB38 nad RAB32. These proteins play a crucial role in the transport of melanogenic enzymes during melanosome maturation in vertebrate melanocytes. Targeted expression of the double mutant in the GTP binding pocket of Ci-Rab38/32/rab-rp1/ltd protein in *Ciona* pigment cell lineage, leads to the formation of larvae with defects or completely lacking of pigment cells in the sensory vesicle.

CHAPTER 1

INTRODUCTION

1.1 Pigment cells: different evolutionary strategies of the same structure

In vertebrates three types of melanin-producing pigment cells are known: melanocytes of the inner ear, skin, hair-bulbs and uvea, which derive from the neural crest; retinal pigment epithelium (RPE) cells of the eye derived from the neural tube; and pigment cells of the pineal organ, which also arise from the neural tube (Eakin, 1973; LeDouarin, 1982; Silvers, 1979). Despite their different embryonic origin, all these cells have in common the production of melanins, that consist of copolymers of black and brown eumelanins and red and yellow pheomelanins (Wakamatsu et al., 2002). These complex pigments are widespread in all prokaryotic and eukaryotic kingdoms, and because of their physical and chemical properties, they have been adopted, during evolution, to cover disparate fundamental functions. Melanin and its intermediates have a large set of manifold purposes that range from increasing cell structural rigidity, as it is in plant seed pods and insect cuticles (Riley, 1992), physiological buffering roles, such as photoreceptor shielding, thermoregulation, photo-protection (King-Smith and Cronin, 1996); (Sarna, 1996); antibiotic and chemo-protective functions (Sichel et al., 1987); electron acceptor functions, in bacteria; increasing virulence factors, in bacteria and fungi (Plonka and Grabacka, 2006; Riley, 1992). In animals, both invertebrate and vertebrate, melanin plays fundamental roles also in inter/intra-specific behaviour (camouflage, sexual display and mating) (Wittkopp et al., 2002). In mammals, melanin serves to minimize the damage caused by exposure to the UV radiation from sunlight, provides color to the skin, hair and eyes and it also functions in the development of the optic nervous system and regulate retinal function (Jeffery, 2001).

1.1.1 Melanin biosynthesis pathway

Melanin pigments, derived mainly from the oxidation of tyrosine (Plonka and Grabacka, 2006), are heterogeneous macromolecules formed by 5,6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-carboxylic acid (DHICA) (brown-black eumelanin), or derived from the sulphur containing cysteinyl-3,4-dihydroxyphenylalanine (cysteinyl-DOPA) (yellow-reddish pheomelanin) (Meredith and Sarna, 2006). Melanin biosynthesis is a multi-step process that comprises several biochemical reactions, mainly using three highly conserved enzymes: Tyrosinase, Tyrosinase Related Protein-1 (TRP-1) and Tyrosinase Related

Protein-2 (TRP-2). All together, they catalyze melanin biosynthesis, using the amino-acid L-tyrosine, as initial substrate. In particular, Tyrosinase is considered the key enzyme of the pathway because, solely, it is responsible for the first two reaction steps,and of the final steps of melanin biosynthesis, that is the oxidation of 5,6-dihydroxyindole (DHI) to indole-5,6-quinone (Fig. 1.1). Furthermore, it also catalyzes The presence of metal ions (such as Zinc and Copper) and TRP-2, also named DOPAchrome tautomerase, bring to conversion of the DOPAchrome (red) in the 5,6-dihydroxyindole-5-carboxylic acid (DHICA) intermediate, rather than in DHI (Palumbo et al., 1991). TRP-1, or DHICAoxidase, is involved in the oxidation of DHICA to indole-5,6-quinone carboxylic acid, one of the final reaction steps (Fig.1.1) (del Marmol and Beermann, 1996).
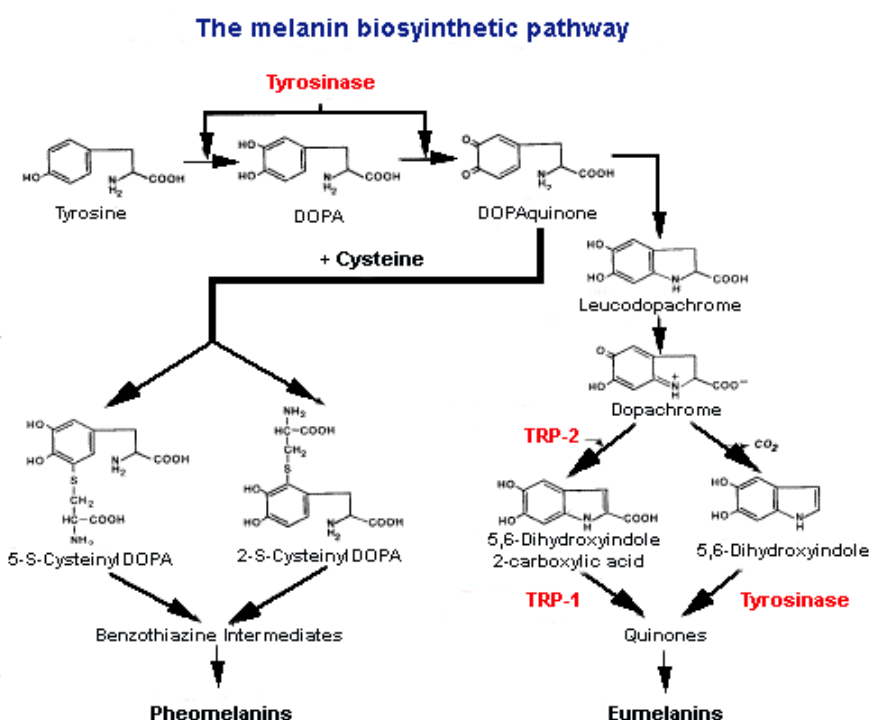
Fig. 1.1 Graphical representation of the melanin biosynthetic pathway. Tyrosinase, the key enzyme of the melanin synthesis, participates in three reactions: hydroxylation of L-tyrosine to 3,4-dihydroxyphenylalanine (DOPA), oxidation of DOPA to DOPAquinone and oxidation of 5,6-dihydroxyindole (DHI) to indole-5,6-quinone. DOPA, 3,4- dihydroxyphenylalanine; TRP-1, Tyrosinase Related Protein-1; TRP-2, Tyrosinase Related Protein-2. Adapted from www.iconocast.com.

In vitro the production of melanin requires the substrate L-tyrosine with the sole presence of tyrosinase, suggesting that melanogenesis process mainly needs this single enzyme. This is also the case in bacteria, sponges and plants. In the course of animal evolution, however, the situation has become more complex up to the mammals where the process is very sophisticated and has to be tightly regulated in terms of amount, type of melanin produced and the environment in which the synthesis takes place. Thus we move from a simple to a complex multi-enzymatic process as in mammalian melanocytes, where new family

members, the TRPs, have been added to finely tune the whole pathway. Sequence comparison of tyrosinase and TYRPs reveals that these proteins share many key structural features (Esposito et al., 2012), indicating their common origin from an ancestral tyrosinase gene able to catalyze the critical rate-limiting hydroxylation of L-tyrosine to L-DOPA (Olivares and Solano, 2009).

1.1.2 Melanin production and storage

Depending on the considered species, melanin production can occur in extra or intracellular environments; for instance, in bacteria melanin is directly synthesized in the surrounding environment, where it acts as antitoxic and chemo-protective agent (Claus and Decker, 2006). In several species of pathogenic fungi, melanin can be produced by different cell types during host infection (Takano et al., 1997). In vertebrates, melanin synthesis occurs in melanosomes that are intracellular organelles present in different vertebrate organs and tissues, where they accomplish various tasks. They are located in Retinal Pigment Epithelium (RPE) and choroidal pigment cells of the eye,(Marks and Seabra, 2001); in the stria vascularis of the inner ear (Steel and Barkway, 1989) and in the basal cell layer of the epidermis, where they are transferred by the dendrites of the epidermal melanocytes into the keratinocytes, residing in the squamous cell layer, (Sulaimon and Kitchell, 2003). Melanosomes are considered to be lysosome-related organelles (LROs), deriving from early endosomal membranes uniquely generated by pigment cells to synthesize and store melanin pigments. Sequestration within melanosomes protects components of the cytosol and other membranous organelles from oxidative attack during melanin synthesis, and concentrates melanins for storage (in eye pigment cells) or cell transfer (from epidermal melanocytes to keratinocytes in the skin and hair) (Raposo and Marks, 2007).

1.2 Vertebrate melanin producing pigment cells

In mammals melanosomes mature within the melanocyte or the developing retinal pigment epithelial cell through four morphologically distinct stages referred to as stages I-IV (Fig 1.2). Stage I melanosomes are characterized by the presence of intraluminal vesicles and fibrils that emanate from the intraluminal membrane. Stage II melanosomes have an oval shape and contain fibril-derived parallel sheet-like structures, on which melanin is deposited at later stages. In stage III, melanogenic enzymes, such as tyrosinase and Tyrp1 (tyrosinase-related protein 1), are transported to premelanosomes and melanin pigments are synthesized and deposited until melanosomes formation is completed in stage IV. In epidermal melanocytes, stage IV melanosomes are translocated along microtubules from

the cell centre to actin-rich dendritic tips for transfer to keratinocytes, in the case of skin melanocytes, or long-term storage, in the case of retinal pigmented epithelial cells in the eye (Ohbayashi and Fukuda, 2012; Wasmeier et al., 2006).
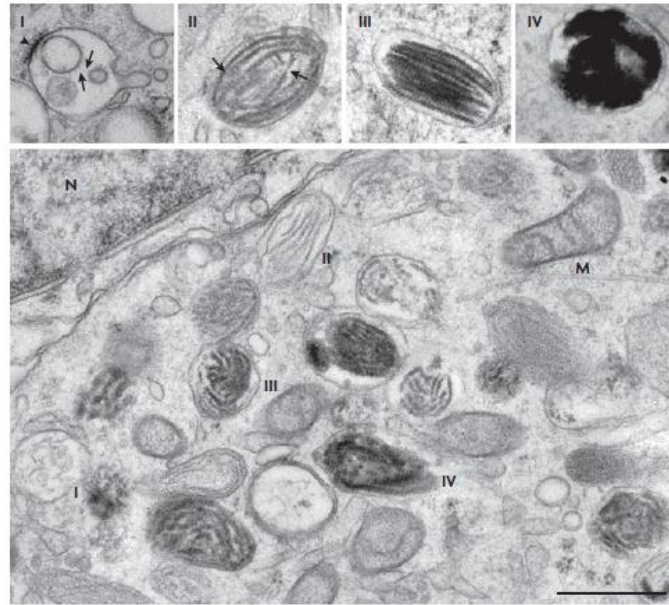
Fig. 1.2 Ultrastructural characterization of melanosomes. Electron microscopy analyses of MNT-1 human melanoma cells fixed by high-pressure freezing before cryosubstitution and embedding. The four stages of melanosome development are shown in the upper panels. Note the dense bilayered coat (arrowhead) and intralumenal vesicles (arrow) of stage I melanosomes, the proteinaceous fibrils (arrow) of stage II, and the melanin deposition (black) in stages III and IV. The main panel shows a typical field of MNT-1 cytoplasm near the nucleus, which contains all four stages of melanosomes. M, mitochondria; N, nucleus. Scale bar represents 0.5 μm.(Adapted from Rapos G. and Marks M.S. 2007).

There is a strong interest to shed light on melanosome development and function not only related to basic research aims but, above all, to biomedical aspects. In fact, many genetic inheritable pathologies, associated to pigment cell gene regulative network, have been reported; such diseases are characterized by reduced or absent pigmentation and are often combined with additional seemingly unrelated disorders such as bleeding diathesis, lung fibrosis and immunodeficiency due to generalized LRO defects (Spritz et al., 2003). Furthermore multiple forms of albinism, vitiligo and deafness are linked to genetic mutations in one or more of the different genes responsible for melanin biosynthesis and deregulations of key genes that rule pigment cell differentiation and proliferation provoke malignant transformation of pigment cells in melanoma skin cancer (Goding, 2007). The enzymatic machinery and structural components of melanosomes are synthesized superficially by pigment cells and must be appropriately targeted to newly forming melanosomes. Best known components are integral membrane proteins that are uniquely expressed in pigment cells, and mutations in the genes that encode them often cause albinism in humans or coat-colour dilution in animals (Bennett and Lamoreux, 2003). Melanosomes also harbour ubiquitous components of lysosomes and late endosomes

(Orlow, 1995). Melanosomes in skin melanocytes coexist with separate bona fide lysosomes (Raposo et al., 2001), a feature that they share with a class of LRO, comprising platelet dense granules, lung type II epithelial-cell lamellar bodies and others (Raposo and Marks, 2007). To generate these LROs, host cells must correctly sort macromolecules that are destined for conventional lysosomes from those destined for the LRO. This sorting is unique to specialized cell types, as shown by the observation that most melanosomal proteins localize to late endosomes and lysosomes when they are expressed ectopically in non-pigment cells (Berson et al., 2001; Bouchard et al., 1989). Melanosomal protein sorting is uniquely defective in the genetic disease known as Hermansky–Pudlak syndrome (HPS), a multisystem disorder characterized by partial albinism, excessive bleeding and often lug fibrosis, sometimes accompanied by immunodeficiency and/or granulomatous colitis (Wei, 2006).

1.3 The Rab GTPases family and their role in Membrane Traffic

The cytoplasm of a typical eukaryotic cell is populated with a variety of membranous organelles, and a vast array of factors traffic between these organelles by vesicular transport. The molecular mechanisms by which this traffic is regulated, to ensure both the fidelity and efficiency of transport has been, and will continue to be a crucial point. Among literature referring to this topic, in the last years several researches have been focused on studying the mechanisms of cargo selection, the budding and scission of vesicles from their donor membranes, the assortment of coats that associate with these vesicles, the mechanism by which these vesicles are transported along cytoskeletal components such as actin filaments or microtubules, the association of the vesicles with the correct target membrane through diverse "tethering" complexes, and finally the mechanism of vesicle fusion with the target membrane through the action of soluble NSF attachment protein receptors (SNAREs that are Soluble N-ethylmaleimidesensitive factor attachment protein receptor) and their associated regulatory machinery. A still open question is how the identity of each organelle is maintained such that their assigned functions and the directionality of transport are not lost as cargo is actively exchanged with other organelles. The Rab GTPases have been identified as key regulatory factors of membrane trafficking. Specific Rabs are physically associated with each organelle as well as their associated transport vesicles. Rabs constitute the largest family of small Ras-like GTPases with 11 identified in yeast and more than 60 members in humans that can be classified in several functional groups. Originally identified as regulators of membrane trafficking in yeast, the Rabs have been found to regulate membrane trafficking in a large number of species

ranging from amoeba and yeast to plants, nematodes, insects, and humans (Pereira-Leal and Seabra, 2001).

1.3.1 Rab proteins as molecular switches

As other monomeric Ras-like GTPases, Rab proteins constitute the largest family of small GTPases, which function as molecular switches that alternate between two conformational states: a GDP-bound "inactive state" and a GTP-bound "active state"; this switching cycle is controlled by two regulatory enzymes, guanine nucleotide exchange factor (GEF) and GTPase-activating protein (GAP) (Fig. 1.3). GEF proteins act on the membrane-inserted Rab to convert it to the GTP-active form that is recruited to transport vesicles/organelles, and it promotes their trafficking (such as vesicle budding, vesicle motility, vesicle docking to specific membranes, and/or vesicle fusion) by interacting with specific effector molecules. GAP proteins bind to the Rab to catalyze hydrolysis of the bound GTP to GDP and thereby convert the Rab back to its inactive state (Pfeffer, 2001). Although there are rare examples of Rab effectors that prefer the GDP-bound form (Shirane and Nakayama, 2006), this form can be considered to be the inactive state of Rab GTPases. The inactive Rab is then a substrate for GDI (GDP dissociation inhibitor) which is able to extract the Rab in its GDP-bound conformation from the membrane. Rab GDI was originally identified as a factor that prevents release of GDP from Rab, thereby stabilizing the inactive form (Matsui et al., 1990). In its GDP-inactive form is subsequently inserted into its respective membrane. The Rab, bound to GDI, is ready to be reinserted into a membrane and begin the cycle again (Fig.1.3).
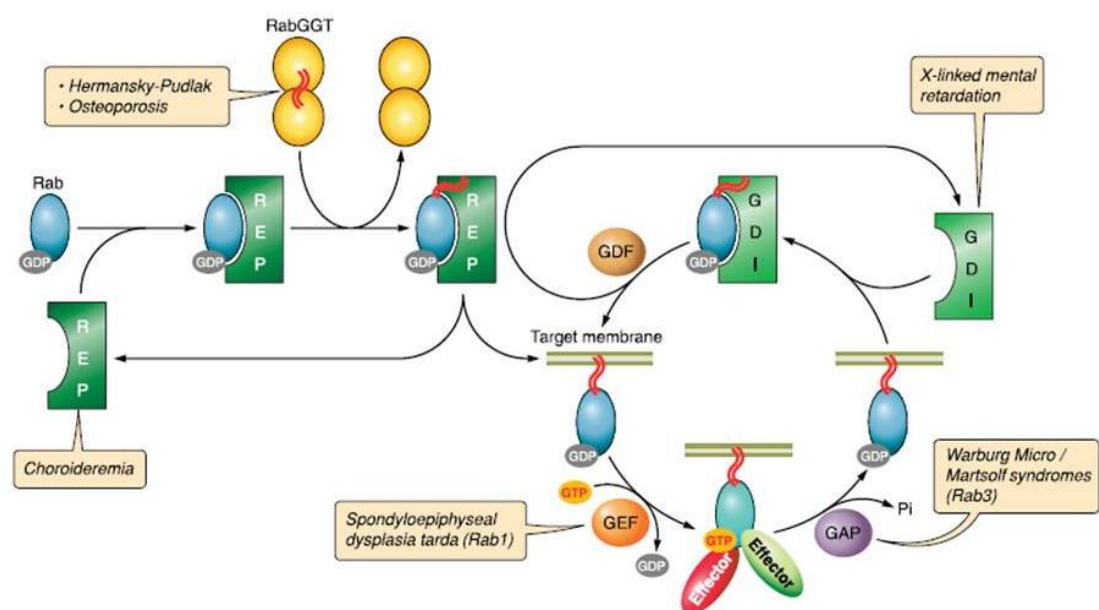
Fig.1.3 The Rab cycle. The newly synthesized Rab protein associates with Rab escort protein (REP) that directs it to Rab geranylgeranyl transferase (RabGGT) to receive its prenyl tails (red wavy lines). REP

delivers the Rab to its target membrane. Throughout this process, the Rab is GDP-bound. A guanine nucleotide exchange factor (GEF) catalyzes exchange of GDP for GTP to activate the Rab. The GTP-bound Rab interacts with effector proteins that mediate membrane traffic in the pathway regulated by its associated Rab. The Rab then interacts with its associated GTPase activating protein (GAP) that catalyzes hydrolysis of GTP to GDP by the Rab. The Rab is then removed from the membrane by guanine nucleotide dissociation inhibitor (GDI) in preparation for the next cycle. The insertion of the Rab into the target membrane is mediated by a GDI dissociation factor (GDF) that releases the Rab from GDI. Loss-of-function mutations at each of the above steps produce disease phenotypes as indicated by the red text boxes. Adapted from (Hutagalung A.H. and Novick P.J.. 2009).

The Rab cycle is critical for regulating traffic to and from particular organelles and thus helps to define their identity. Any perturbation in the steps described above can result in a variety of disease states (Fig. 1.3). Mutations in the human REP-1 gene lead to choroideremia, a disease characterized by progressive atrophy of the choroid, retinal pigment epithelium, and retina that lead to eventual blindness; the cause of the disease is most likely due to loss of Rab27A function, which accumulates in an unprenylated form in retinal tissue samples from patients with the disease (Matesic et al., 2001). Modulation of RabGGT function has also been shown to play a role in several diseases. The mouse *gunmetal* mutant is a RabGGT α loss-of-function mutant that is phenotypically similar to patients with Hermansky-Pudlak syndrome (Bonifacino, 2004; Di Pietro and Dell'Angelica, 2005). Mutations in the human GDI1 gene lead to X-linked nonspecific mental retardation (D'Adamo et al., 1998). Warburg Micro and Martsolf syndromes are associated to mutations in the genes encoding the regulatory and catalytic subunits of the Rab3GAP; these diseases are characterized by developmental abnormalities of the eye, nervous system, and genitalia (Aligianis et al., 2005).

1.3.2 Structural mechanisms for regulation of the Rab GTPase cycle

Rabs have been classified in several phylogenetic and functional groups and at least for a crystal structure is available, allowing for some generalization regarding the specific structural features that contribute to Rab function (Pfeffer, 2005). Rabs generally possess the GTPase fold, composed of a six-stranded β-sheet flanked by five α-helices, common to all members of the Ras superfamily (Fig. 1.4 A). Rab GTPases are reversibly associated with membranes by hydrophobic geranylgeranyl groups that are attached to one or (in most cases) two carboxy-terminal Cys residues, and this is intrinsic to their role in regulating membrane traffic. Because of the overall structural conservation, the differences between the active and inactive states must define the regions that determine the specific functions of each Rab. Rab proteins possess a characteristic GTPase fold with two 'switch regions' that contact the γ phosphate of GTP and exhibit large conformational differences between the inactive and active states (Fig 1.4 B-C). An essential Mg^{2+} cofactor is required for high

affinity nucleotide binding and hydrolysis. Interactions with guanine nucleotides involve mainly residues from five 'G motifs' that are broadly conserved in GTPases (Vetter and Wittinghofer, 2001). The GxxxxGK(S/T) motif in the 'P-loop' ($\beta 1/\alpha 1$ loop) provides phosphate contacts and supplies a Ser/Thr that is co-ordinated by the Mg^{2+} ion. A threonine in switch I contacts the γ phosphate and is co-ordinated by the Mg^{2+} ion in the active state. The DxxGQ motif in switch II provides an aspartate that stabilizes the Mg^{2+} ion through interaction with a water ligand, a glycine that contacts the γ phosphate, and a glutamine that functions as a catalytic residue for the intrinsic GTP hydrolytic reaction.
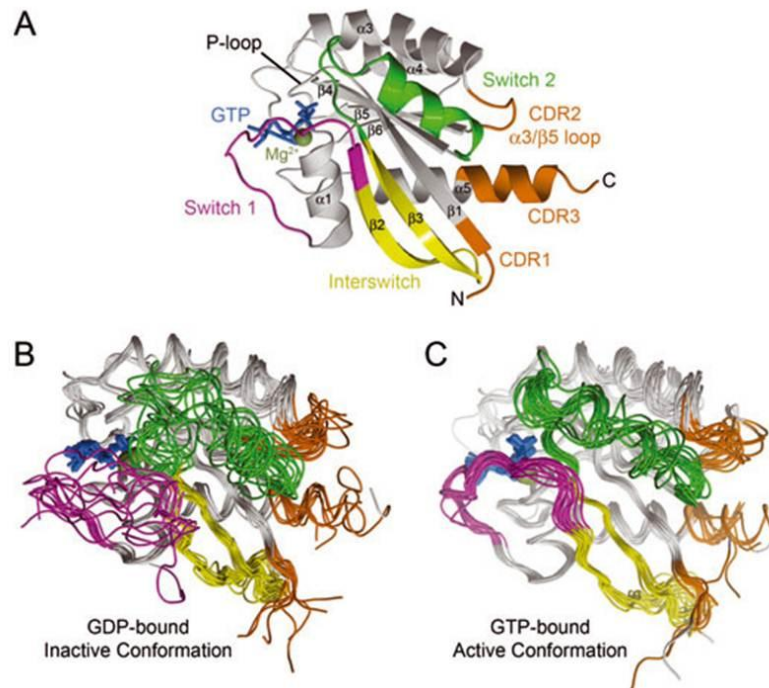
Fig. 1.4 Structural similarity and diversity within the Rab family. A) Overall structure of a representative Rab GTPase (Rab3) with functional regions colored as indicated. B) Comparison of inactive GDP-bound Rab GTPase structures after superposition with Rab2. Note that the switch regions are either poorly ordered or adopt ordered conformations as a result of crystal packing. C) Comparison of active GppNHp-bound Rab GTPase structures after superposition with Rab3. Although both switch regions adopt stable active conformations, switch II exhibits large conformational differences betweenRab GTPases. Adapted from (Lee M, et al., 2009).

1.3.3 Rab proteins in melanosomes biogenesis

The melanosomes formation has been heavily studied for disease implications caused by defects in melanosomes formation and because the melanosomes is a prototype of the specialized class of organelles called lysosome-related organelles (LROs). During melanosomes formation, tyrosinase and tyrosinase-related proteins-1 and -2 (Tyrp1 and Tyrp2), the main proteins responsible for melanin biosynthesis are transported to the maturing melanosome requiring a sorting step at specialized tubular domains of early/recycling endosomes, rather than direct transport from the trans-Golgi network. Packaging of the tyrosinases into transport vesicles at early/recycling endosome-associated

tubules is dependent on ubiquitous adaptor protein complex (AP)-1 and AP-3, and biogenesis of lysosome related organelles complex (BLOC)-1 and BLOC-2.16,17 Furthermore, AP-3 and BLOC-2 define parallel pathways for melanosome biogenesis, thus, deficiency of both complexes causes a more severe defect in Tyrp1 transport and overall pigmentation than either single deficiency (Fig. 1.5).
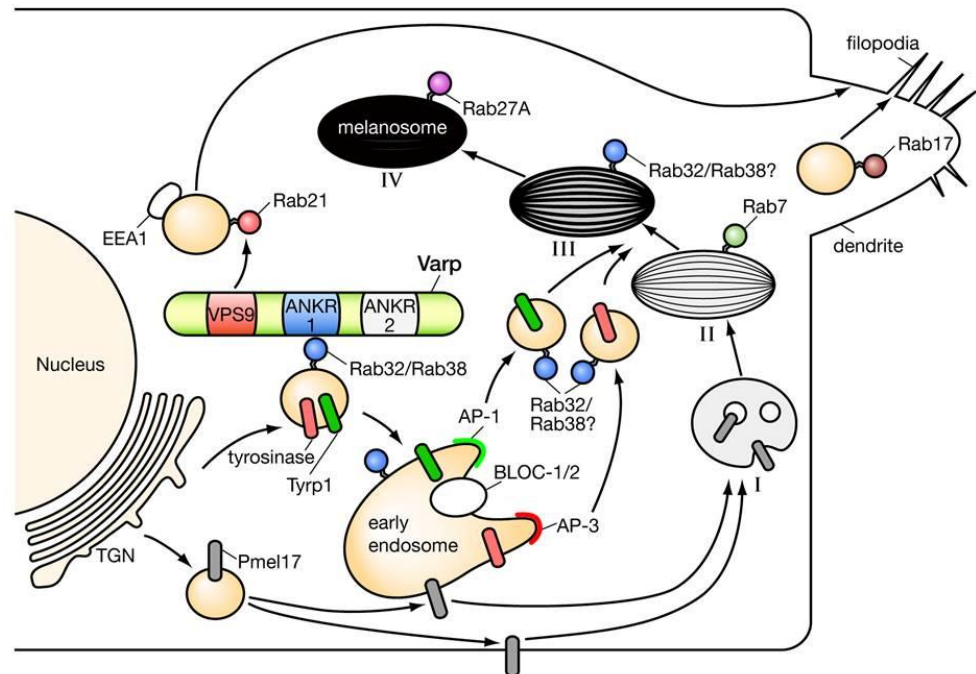
Fig. 1.5 Schematic diagram of endosomal transport systems that regulate melanosome biogenesis (from stage I to IV) and dendrite formation. The endosomal transport systems play important roles in the transport of tyrosinase and Tyrp1, because these melanogenic enzymes are transported from the TGN to stage III melanosomes via early endosomes. The BLOC and AP complexes are key players in the endosomaltransport systems . AP-3 is required for tyrosinase transport from early endosomes to melanosomes, whereas BLOC-1/2 and AP-1 are required for Tyrp1 transport. Rab32/Rab38 and Varp are additional factors that are required for the transport of melanogenic enzymes, presumably for a step different from that regulated by BLOCs/APs. Varp also functions as a GEF for Rab21 to promote EEA1/Rab21-positive endosome trafficking for dendrite formation , which is essential for efficient melanosome transfer to neighbouring keratinocytes. Rab17promotes the formation of dendritic filopodia by a largely unknown mechanism. Adapted from (Ohbayashi N. and Fukuda M., 2012).

One interesting point is how specialized cells such as the melanocyte simultaneously produce and maintain lysosomes and specialized LROs such as the melanosome. One possible mechanism to facilitate the production and existence of both organelles is the expression of cell type-specific proteins that function in LRO biogenesis and not in lysosome biogenesis. Two closely related small GTPases of the Rab family, Rab32 andRab38, are expressed in selected cell types such as melanocytes**.** Rab32 and Rab38 (Fukuda M., 2008) are localized on tyrosinase- and Tyrp1-containing vesicles and/or organelles around the trans-Golgi network (TGN) as well as on melanosomes (Wasmeier et al., 2006). The weakly diluted mouse coat color mutant *chocolate* has been shown to be caused by dysfunction of Rab38, and targeting of tyrosinase and Tyrp1 to melanosomes is

mildly impaired in *chocolate* epidermal melanocytes (Loftus et al., 2002; Wasmeier et al., 2006). Rab38 and its close homologue Rab32 function redundantly in melanocytes, because additional depletion of Rab32 in cultured chocolate epidermal melanocytes severely impairs the transport of tyrosinase and Tyrp1 to melanosomes, resulting in severer hypopigmentation (Wasmeier et al., 2006). In contrast to its role in epidermal melanocytes, Rab38 has a dominant role in the transport of tyrosinase to immature melanosomes in retinal pigment epithelial cells, because severe impairment of melanosome maturation is observed in the retinal pigment epithelial cells of *chocolate* mice (Lopes et al., 2007). In other organisms, the *Rab38* gene is mutated in *ruby* rats (Oiso et al., 2004), which also exhibit hypopigmentation in addition to the platelet storage pool deficiency related to inherited disorder, such as Hermansky-Pudlak syndrome (HPS) (Wei, 2006), thus *ruby* rats are considered a rat model of HPS. Furthermore, a *D. melanogaster* Rab32/38 ortholog, Rab-RP1, is localized on pigment granules in the eye, and its mutation causes eye color defects (Fujikawa et al., 2002a; Ma, 2004). Although the precise mechanism(s) underlying Rab32/38-mediated transport of melanogenic enzymes is not fully understood, recent identification of a Rab32/38- specific binding protein Varp (VPS9-ankyrin-repeat protein; also called Ankrd27) has greatly improved our understanding of the function of Rab32/38 in melanocytes (Tamura et al., 2009). Varp and Rab32/38 are well colocalized on Tyrp1-containing vesicles in melanocytes and knockdown of Varp or expression of the ANKR1 domain, namely the Rab32/38-binding site, in melanocytes causes a dramatic reduction in Tyrp1 signals on peripheral melanosomes (Tamura et al., 2011).

After melanosomes mature around the nucleus of melanocytes, they are transported from the perinucleus to the dendrite tips along microtubules and actin filaments by the coordinated actions of motor proteins. The rapid, bidirectional microtubule-based movements of melanosomes are regulated by two classes of motors, kinesins and dynein. Following the transfer of melanosomes from microtubules to actin filaments, an actin-based motor myosin Va regulates their slow and unidirectional movement in melanocytes (Fukuda, 2005). The first key regulator of actin-based melanosome transport discovered was myosin Va, and the other two were Rab27A and Slac2-a (also called melanophilin) (Fig. 1.7). The importance of this complex has also been demonstrated by the existence of mutations of the human MYO5A, RAB27A, and SLAC2-A/MLPH genes, which are responsible for human hereditary diseases characterized by silvery hair that are called Griscelli syndrome type I, type II, and type III, respectively(Van Gele et al., 2009). The GTP-bound activated form of Rab27A first binds to the surface of mature melanosomes via the C-terminal geranylgeranylation site and recruits the melanocyte- specific Rab27A effector Slac2-a via the switch II region (Kukimoto-Niino et al., 2008). The resulting

Rab27·Slac2-a complex functions as a cargo (or melanosome) receptor for exon F-containing myosin-Va, which is abundantly expressed in melanocytes (Hume et al., 2006; Wu et al., 2002). Slac2-a is an important linker molecule in this tripartite complex, and the N-terminal SHD (Slp homology domain) and the middle domain (MBD, myosin Va-binding domain) are responsible for Rab27A-binding and myosin Va-binding, respectively (Fig. 1.6) (Ohbayashi and Fukuda, 2012).
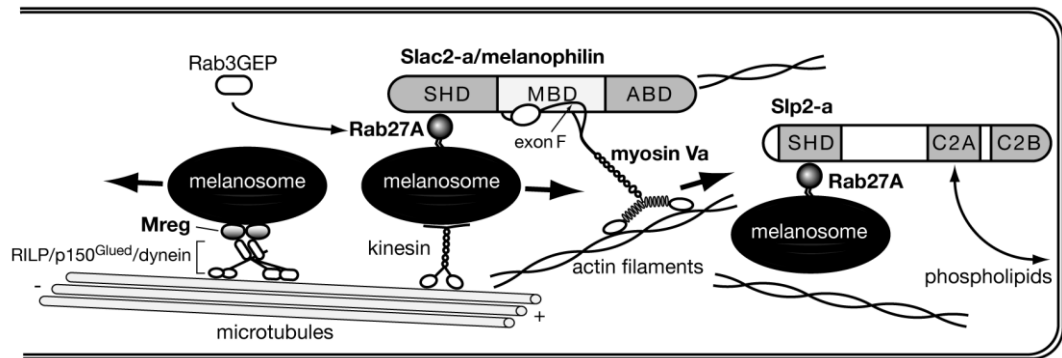
Fig.1.6 Schematic diagram of melanosome transport along microtubules and actin filaments. Mature melanosomes are transported toward the peripheral region via microtubule-based motors (kinesin and dynein), but, except for the discovery that retrograde melanosome transport is regulated by the Mreg-RILP-p150Glued complex in Rab27A-deficient melanocytes, the precise mechanisms underlying the microtubule-based melanosome transport remain to be elucidated. Mreg directly binds the C-terminal domain of RILP and functions as a melanosome receptor for the dynein-dynactin retrograde motor complex. After being transferred to actin filaments, melanosomes are transported by the tripartite complex consisting of Rab27A-Slac2-A-myosin Va and then anchored to the plasma membrane by the Rab27A-Slp2-a complex. Activation of Rab27A by Rab3GEP is required for the recruitment of Rab27A to melanosomes before the actin-based melanosome transport. Adapted from (Ohbayashi N. and Fukuda M., 2012).

1.4 Animal model system: the ascidian *Ciona intestinalis*

To elucidate gene networks' conservation, in animal development, the tunicates are one of the most intensely studied model systems. Tunicates belong to the Urochordata subphylum that, together with Cephalochordata (Lancelets) and Craniata (Myxini and Vertebrates) subphyla, constitute the Chordata phylum (Fig 1.7 A). The Urochordata subphylum is divided into three classes: Ascidiacea (sea squirts), Thaliacea (salps) and Appendicularia or Larvacea (larvaceans). As their phylogenetic position suggests, ascidians tadpole possess a simplified chordate body plan with a dorsal hollow neural tube and notochord (Fig 1.7 B). In contrast to those apparent similarities, the ascidian larval body is strikingly simple compared with that of vertebrate tadpoles. The ascidians larvae are typical swimming tadpole larvae, after substrate settlement, they undergo a process of metamorphosis, becoming adult sessile filter feeder animals (Fig. 1.7 C).
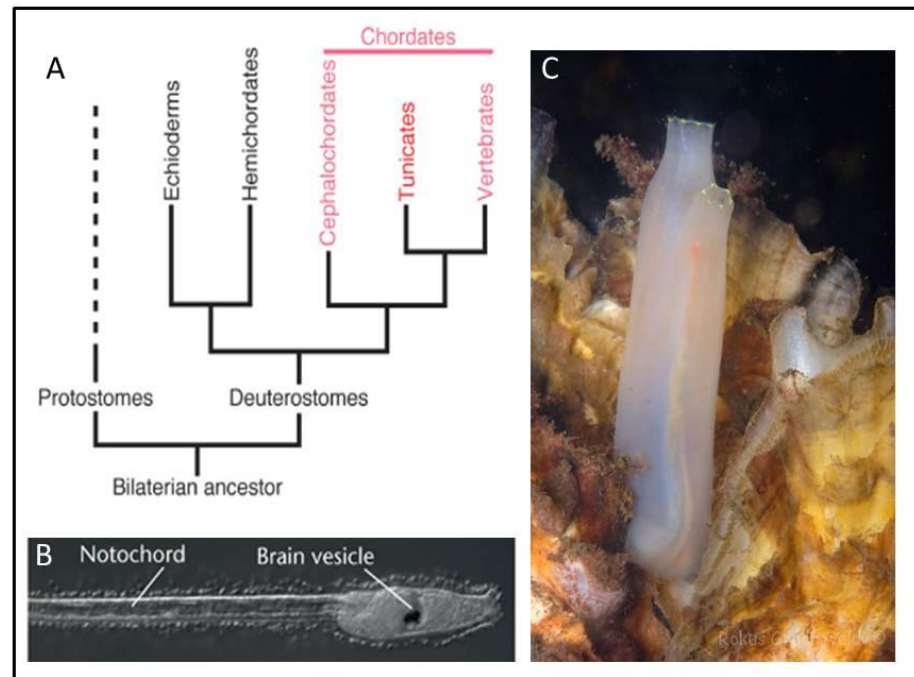
Fig.1.7 The ascidian _Ciona intestinalis_. (A) Phylogenetic relationships of chordates. Ascidians are included in the subphylum Tunicata. B) A _Ciona intestinalis_ larva. The Notochord and the brain sensory vesicle are evidenced. Scale bar: 100 μm. (C) _Ciona intestinalis_ adult. After metamorphosis, _Ciona_ loses its tail and starts to settle.

During metamorphosis, the larval tail is lost and adult tissues grow rapidly, which include characteristic chordate structures such as pharyngeal gills and an endostyle (the endostyle is homologous to vertebrate thyroid gland) (Ogasawara et al., 1999). Traditional studies tend to consider cephalochordates as the closest living relatives of vertebrates, placing tunicates in an earlier evolutionary position. These considerations are mainly supported by morphological similarities in the body plan shared by cephalochordates and vertebrates (Beaster-Jones et al., 2008; Schubert et al., 2006). New phylogenetic analyses, supported the idea that Cephalochordates seem to be the most basal extant chordate while Tunicates are the closest living invertebrate relatives to the Vertebrate (Delsuc et al., 2006; Vienne and Pontarotti, 2006). Concerning morphological divergence, they formulated an explanation, hypothesizing that the aforementioned vertebrate-like structures should have been ancestral features of the common chordate ancestor. They might have been retained in relatively slowly-evolving organisms, as cephalochordates and vertebrates; instead, they might have been secondarily lost in tunicates, due to their faster genome evolution rate. This point of view might revaluate tunicates as highly diverged and specialized organisms, rather than primitive chordates; nonetheless, the debate is still open in the scientific community.

1.4.1 Characteristics of Ciona intestinalis as an experimental system for genetics

Ciona intestinalis is a cosmopolitan ascidian, available during most part of the year; adults can be bred in laboratory facilities permitting easily the collection of gametes and embryos to work with. An adult can bear thousands of eggs; gametes can be collected surgically and eggs can be fertilized with sperm from different individuals. They subsequently exhibit normal, synchronized development. *Ciona* embryos show a relatively fast developmental time from the zygote to the tadpole larval stage(~16 hours post fertilization at 18°C) (Fig.1.9) (Hotta et al., 2007).

Since they are characterized by an invariant, well known, embryological cleavage program, starting from the 20th Century, they have been object of studies on cellular lineage and on maternal determinant factors involved in the specification of embryonic territories, during the first hours of development (Jeffery and Meier, 1983; Zalokar and Sardet, 1984). Embryo manipulations are very easy to perform; transgenic experiments in *C. intestinalis* are facilitated by the introduction of plasmid DNAs into the fertilized eggs by a simple electroporation technique (Corbo et al., 1997). This method allows rapid screening of promoter fragments fused to a reporter gene such as Green Fluorescent Protein (GFP) to identify regulatory elements controlling gene expression; furthermore in ascidians every blastomere of the embryo is distinguishable up to neurula stage, so that it is easy to precisely identify cells expressing genes of interest, when gene expression is initiated and lineage in which gene expression is inherited (Satoh, 2001). Generally, tunicate genomes are quite compact with relatively short regulatory regions in close proximity to the gene coding sequences. This peculiarity coupled with the possibility, in *C. intestinalis*, to easily isolate the promoter regions of the gene of interest, (Di Gregorio and Levine, 2002), makes *Ciona* a model system ideal to identify marker genes, specific for each lineage, and study the genetic cascades in which they are involved. Genomic and cDNA libraries have been produced and are available for the tunicate research community (Dehal et al., 2002) Furthermore Affymetrix GeneChip microarrays have been custom designed for *C. intestinalis* genome (Christiaen et al., 2008) resulting as a good technique for gene expression analysis.

1.4.2 Pigment cells in the Ciona intestinalis Central Nervous System

Recent studies reveal that ascidian larva and vertebrate embryo show extensive evolutionary and molecular homologies. The ascidian larva has a miniature CNS with many chordate credentials. It contains less than 100 neurons and 250 glial cells whilst the peripheral nervous system is composed of 20-30 epidermal sensory neurons (Meinertzhagen and Okamura, 2001). The CNS is conventionally divided, from the

anterior to the posterior part of the larval body, into four regions: an anterior sensory vesicle (SV), connected by a neck to a visceral ganglion, and a dorsal caudal nerve cord. The sensory vesicle is a hollow cavity sited in the trunk of the larva that could be morphologically divided into three regions: the anterior SV, middle SV and posterior SV (Fig 1.8).
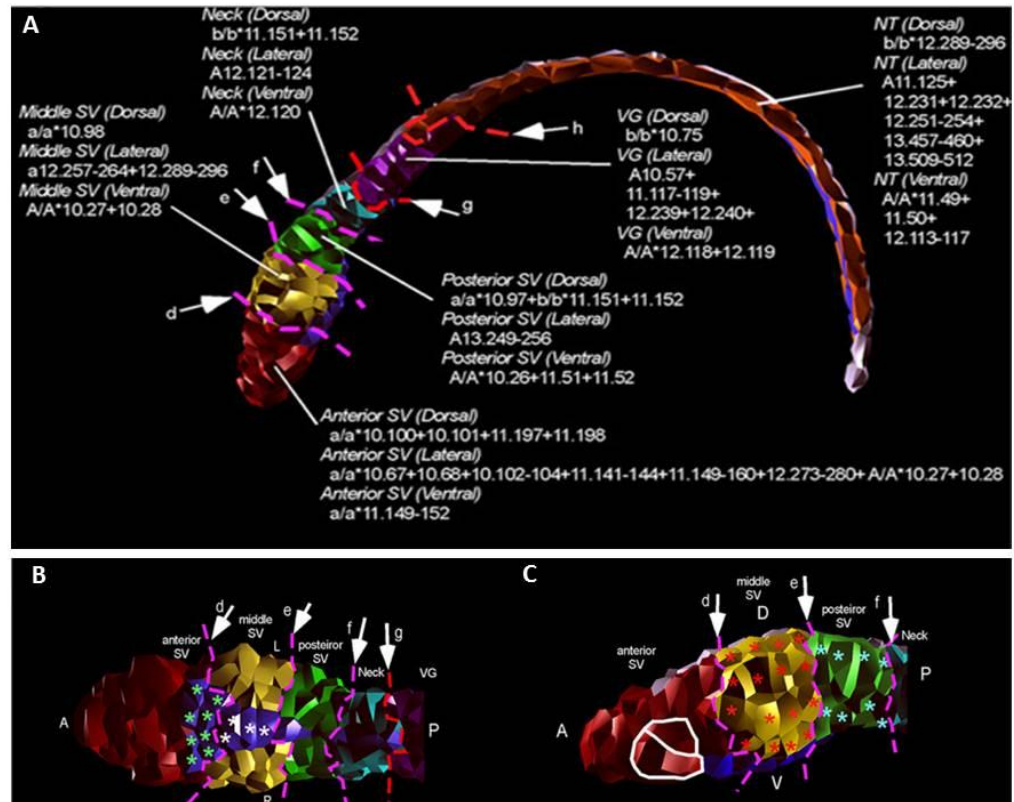
Fig 1.8 Central Nervous System (CNS) cells at larval stage visualized in the 3DVMTE: (A) Lateral side cells in six regions: anterior Sensory Vesiscle (SV) (red), middle SV (yellow), posterior SV (green), neck (lightblue), Visceral Ganglion (VG) (purple) and caudal Neural Tube (NT)(orange) of the CNS. Dorsal (white) and ventral (blue) cells are also shown. Pink (arrows d, e, and f) and red (arrows g and h) dotted lines represent the boundaries of each region. (B) Ventral view of the anterior SV, middle SV, posterior SV, neck, and anterior VG regions. Cells of the ventral rows of the anterior SV and middle SV are indicated by green and white asterisks, respectively. (C) Lateral view of anterior SV, middle SV, posterior SV, and anterior neck regions. Lateral cells in the middle SV and posterior SV are indicated by red and light blue asterisks, respectively. Two cells outlined in white are large cells found on both lateral sides of the anterior SV. These cells are larger than all neighbouring anterior SV cells and are elongated along the anterior-posterior axis.

On the basis of morphology and expression of marker genes, the sensory vesiscle has been considered the homolog of the vertebrate prosencephalon and mesencephalon (Wada et al., 1998). In the anterior sensory vesiscle two pigment cell-containing sensory structures are present: the anterior geotactic otolith and the posterior photoreceptive ocellus (Fig. 1.9) (Dilly, 1969; Eakin and Kuda, 1971; Ohtsuki, 1991; Torrence, 1986; Tsuda et al., 2003a), which directs larval swimming behaviour before metamorphosis (Tsuda et al., 2003b).
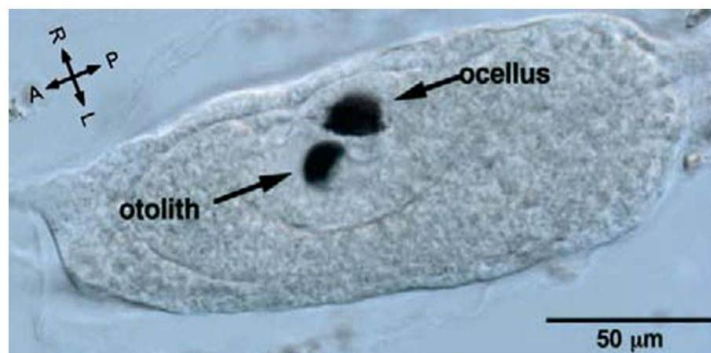
Fig. 1.9 Larval sensory vesicle. The black arrows indicate the two pigment cell sensory organs in the larval sensory vesicle, the otolith and the ocellus. Adapted from (Tsuda et al., 2003)

The otolith is formed by a single, highly specialized, peculiar cell. Apart from a nucleus, the body of this cell, the statocyte, contains a large, round-shaped melanin granule within a vacuole. The statocyte is connected by a narrow stalk to the ventral floor of the sensory vesicle. It has been shown that the functional role of this organ is gravity perception(Tsuda et al., 2003b). The ocellus shows it is a multicellular organ, composed by three lens cells, about 30 photoreceptor cells in *C. intestinalis* and a pigment cell. Photoreceptor cells are sited in the right, posterior wall of the sensory vesicle (Horie et al., 2005). The ocellus pigment cell is cup shaped and, contains several membrane-bound pigment granules. In the bottom part of the ocellus structure, between photoreceptors and pigment cell, the three lens cells are located (Fig. 1.10).
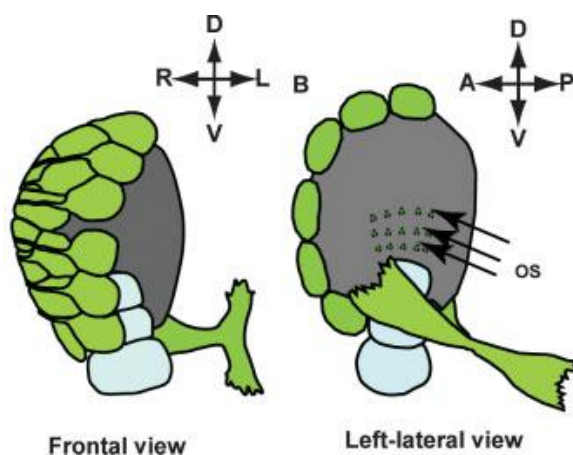
Fig. 1.10 Ocellus graphical representation. The ocellus sensory organ is formed by three parts: photoreceptor cells (green), pigment cell (gray), lens cells (pale blue). A: Frontal view. B: Left-lateral view. OS: photoreceptors' outer segments. Adapted from (Horie et al., 2005).

The ocellus sensory organ is involved in larval photoreception. In particular, pigment cell melanin granules filter directionally the entering light and play a photoprotective role for the posterior photoreceptors (Tsuda et al., 2003a). Functional similarity and conservation

of basic gene expression patterns brought some authors to infer that tunicate sensory organs, vertebrate eye and pineal organ could have been derived from a common archetypal "visual organ" (Kusakabe et al., 2001; Lamb et al., 2007; Sato and Yamamoto, 2001).

Molecules belonging to arrestin and opsin classes are expressed in ocellus photoreceptors as in vertebrate retina photoreceptor cells and pineal organ (Kusakabe et al., 2001; Nakagawa et al., 2002; Nakashima et al., 2003). Ascidian sensory organ pigment cells, analogously to the vertebrate RPE, pineal gland and melanocytes (Tief et al., 1996), express the melanogenic genes tyrosinase (tyr) (Caracciolo et al., 1997; Toyoda et al., 2004) and tyrosinase related protein (tyrp) (Toyoda et al., 2004). Moreover in both ascidians and vertebrates, these genes can be considered as early developmental pigment cell markers, as their transcript production is not restricted to differentiated pigment cells, but appears in the early pigment cell precursor lineage (Caracciolo et al., 1997; del Marmol and Beermann, 1996; Palumbo et al., 1991; Sato and Yamamoto, 2001; Tief et al., 1996). Recently, it has been demonstrated that Tyrosinase family members, *Ci-tyr*, *Ci-tyrp1/2a* and *Ci-tyrp1/2b*, are expressed from the late gastrula stage up to the tadpole larval stage in pigment cell lineage (Fig. 1.11 A-L) and they can considered as specific marker genes for pigmented cell precursors in *Ciona intestinalis* (Esposito et al., 2012).
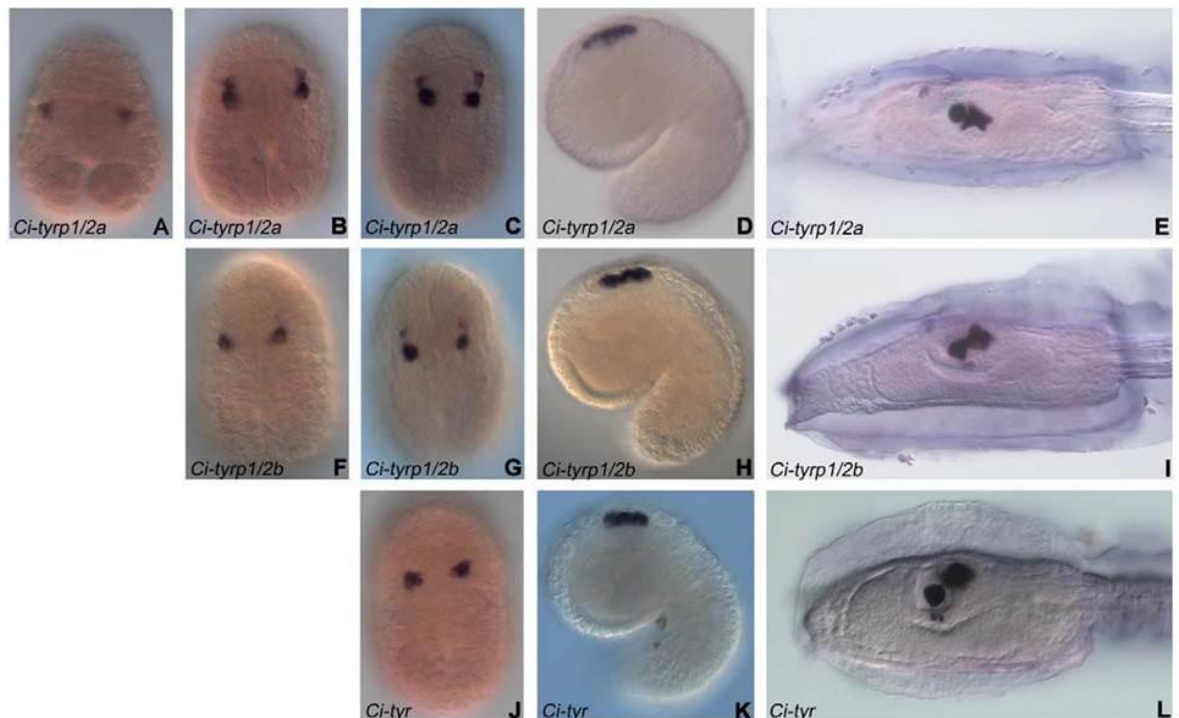
Fig. 1.11 Ci-tyr-tyrps expression in pigment cell precursors. Expression pattern of Ci-tyrp1/2a (A–E), Ci-tyrp1/2b (F–I) and Ci-tyr (J–L) in C.intestinalis embryos at different developmental stages, detected through whole mount in situ hybridization experiments. The three genes are specifically expressed in a9.49 blastomeres (pigment cell precursors) and in their descendant cells from late gastrula (Ci-tyrp1/2a, in A), middle neurula(Ci-tyrp1/2b, in F) or late neurula (Ci-tyr, in J) stages up to the larval stage (E, I, L). A: late gastrula stage, vegetal view. B, F: middle neurula stage, dorsal view. C, G, J: late neurula stage, dorsal view.

D, H, K: mid-tailbud stage, lateral view. E, I, L: larval stage, lateral view. A, B, C, F, G, J: anterior is up; D, E, H, I, K, L: anterior is on the left. Adapted from (Esposito R., et al., 2012).

Based on the expression profiles in *Ciona*, *tyrp1/2a* appears first, at the late gastrula stage, and is copiously expressed up to the tadpole stage in the pigment cell lineage. This could be in agreement with a need to accumulate a large amount of tyrp protein in order to exert its protective and scaffolding function toward tyrosinase in the melanogenic complex. A further interesting feature, that ascidians share with vertebrates, is that *tyrosinase* transcripts (Fig. 1.11 J-L) and also the corresponding protein product (Whittaker, 1973) appear well before pigment synthesis begins (Steel et al., 1992; Whittaker, 1973). This could be related to the need of a large quantity of enzyme requested at the time of initial melanin synthesis, in order to catalyze the production of sufficient L-DOPA cofactor to maintain a rapid and sustained tyrosine oxidase activity (Esposito et al., 2012). The ascidian adult specimens have a sessile lifestyle and during metamorphosis to the adult stage both the two pigment cell sensory organs undergo the process of apoptosis and reabsorption, as the majority of the larval body structures; hence, these pigmented sensory organs are necessary only in the first part of the ascidian's life.

1.4.3 FGF/MAPK/Ets signaling has a fundamental in the induction of pigment cells in the sensory organs of C. intestinalis

Recently it has been demonstrated that the FGF signaling pathway is the inductive signal responsible for conditional pigmented cell precursors specification (Squarzoni et al., 2011). *C. intestinalis* pigment cell precursors arise at the early gastrula stage (the 110-cell stage) as two symmetric progeny of the a4.2 brain/sensory vesicle lineage, the a8.25 cell pairs (Nishida, 1987). This inductive process lasts for all the gastrulating period, during which the a8.25s divide one time into the a9.49 and a9.50 cell pairs: the a9.50s will participate to form the dorsal and anterior part of sensory vesicle while the a9.49 cell pair are committed as pigment cell precursors by the inductive signal derived from the sub-adjacent A-line nerve cord precursors. At the late gastrula stage, the *C. intestinalis* neural plate is well organized in a grid-like fashion with eight columns and six rows, in which cells are located in invariant positions. The posterior neural plate, comprising rows I and II, derives from the anterior vegetal lineage (denoted as A-line) and contribute to the caudal nerve cord, motor ganglion and posterior sensory vesicle (Imai et al., 2009). Cells in rows III-VI derive from the anterior ectodermal lineage known as the a-line. Progeny of row III contribute to the anterior sensory vesicle, whereas cells of row IV contribute to the stomodeum as well as the anterior brain. Rows V and VI comprise the anteriormost neural plate, and the

progeny of these cells contribute to the palps as well as to epidermal sensory neurons in the trunk.

In this context, the pigment cell precursors (a9.49 cell pair of the neural plate) are in close proximity to two sources of FGF diffusible signaling molecules, Ci-fgf8/17/18 and Ci-fgf9/16/20, in the six A-line nerve cord precursors of row II while Ci-ERK1/2, a known downstream FGF-MAP kinase cascade factor, is active in the adjacent six a-line cells of the neural plate row III, which includes the two pigment cell precursors (Hudson et al., 2007a). During neurulation, the last division of the a9.49 pairs gives rise to two pairs of cells: the a10.98 blastomeres, which form part of the right-dorsal sensory vesicle epithelium and the a.10.97 blastomeres, which become pigment cells (Nishida and Satoh, 1989) (Fig. 1.12).
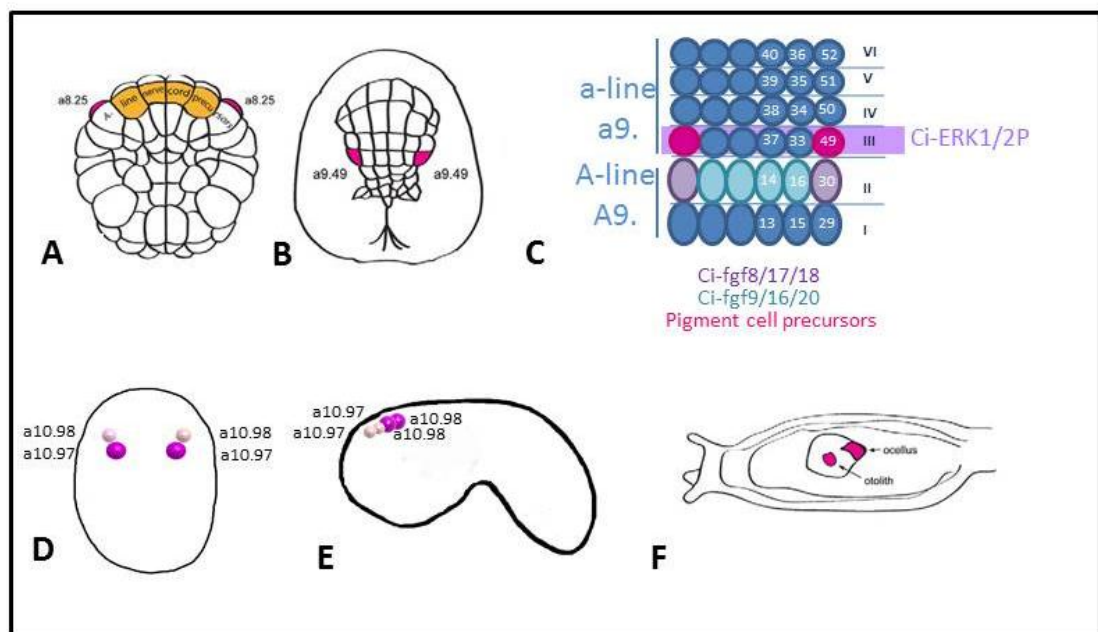
Fig. 1.12 Pigment cell lineage during *C. intestinalis* embryogenesis. During gastrulation (A-D), pigment cell precursor (pink) induction occurs. In this period, A-line nerve cord cells express *Ci-fgf9/16/20* (from 110-cell stage, green) and *Cifgf8/17/18* (from the late gastrula, violet) and *Ci*-ERK1/2 activation appears in row III of the neural plate (C, violet rectangle). (A)☐110-cell stage; (B) late gastrula stage; (C) neural plate scheme at the late gastrula stage; (D) neurula stage (adapted, with permission, from Hudson et al. (Hudson et al., 2007). Vegetal view, anterior is pwards. (E-F)☐During neural tube closure, pigment cell precursors divide and differentiate into sensory organ pigment cells. D, tailbud stage; E, larval stage. Dorsal is upwards, anterior is towards the left.

The FGF signal is transmitted in the cells through the MAPK cascade, which in turn activates the transcription factor Ci-Ets1/2; the expression of *Ci-Ets1/2* was revealed at the late gastrula, early and late neurula stages; during this developmental time period, the transcript appears in a-line neural territories, including pigment cell precursors (Fig. 1.13 A-C); the presence of Ci-Ets1/2 at the right time and place is crucial to behave as FGF signaling effector, serving to instruct the fate of the a9.49 pair down the pigment cell precursor lineage.
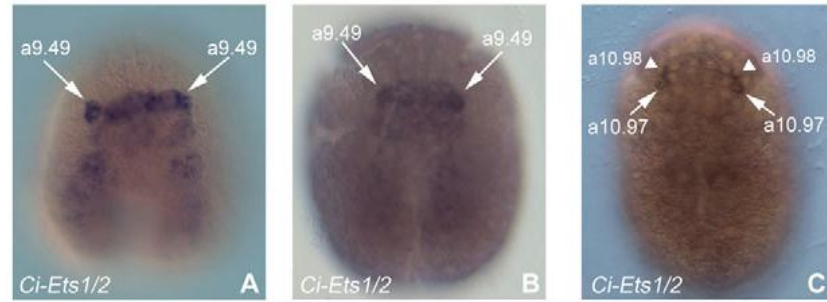
Fig. 1.13 *Ci-Ets 1/2* **expression includes also pigment cell precursors**: (A-C) During gastrulation, *Ci-Ets1/2* is broadly expressed in the neural plate region, including pigment cell lineage (a9.49 in A,B; a10.97 and a10.98 in C). (A)Late gastrula, (B) early neurula and (C) late neurula stages. Vegetal view, anterior is towards the top. Adapted from (Squarzoni P., et al., 2011).

With next cellular division, *Ci-Ets1/2* is partitioned between the two a9.49 cell-progeny, of which the a10.97 pair remains physically close to FGF-emitting sources, *Ci-fgf8/17/18* and *Ci-fgf9/16/20*; considering that FGF molecules act at a very short range from the target cells (Tassy et al., 2006), this strict contact would retain Ci-Ets1/2 activation only in the a10.97 blastomeres, leading to initiation of the pigment cell-specific program; the distance between the source and the target cell is a fundamental requirement for the induction mediated by FGF signaling. Intriguingly, morphogenetic rearrangements that build the *C. intestinalis* embryo morphology push the a10.98 blastomeres in a more rostral position, outdistancing them from FGF sources. In contrast, the a10.97 cell pair remains within the FGF signaling range.

To drive the conditional expression of a dominant-negative form of the sole Ci-FGFR (dnFGFR) in the pigment cell precursors, the promoter region of tyrp1/2a, ptyrp1/2a, was used (Davidson et al., 2006). Targeted expression of dnFGFR resulted in larvae lacking one or both pigment cells, showing normal morphology of the palps, tail and sensory vesicle (Fig. 1.14).
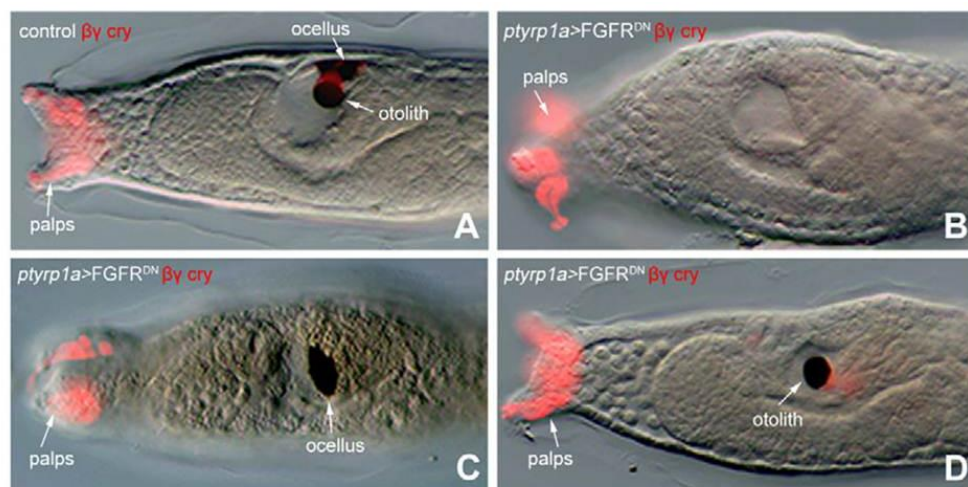
Fig. 1.14   FGF signaling is required for pigment cell induction. (A) Control larvae showing Ci-βγcrystallin expression in the palps and in the otolith. (B-D) All ptyrp1a>dnFGFR-transgenic larvae show Ci-

βγcrystallin expression in the anterior palps. Among larvae possessing one pigment cell (C,D), around 50% show Ci-βγcrystallin in the single otolith-like pigment cell (D). Lateral view, anterior towards the left. Adapted from (Squarzoni P., et al., 2011).

To test the role of *Ci-Ets1/2* in pigment cell precursor induction, constitutively inactive (Ets:WRPW) and active (Ets:VP16) forms of the Ci-Ets1/2 protein have been selectively expressed in pigment cell lineage driven by the ptyrp1/2a regulatory region. The Ets:WRPW repressor form (Davidson et al., 2006; Fisher et al., 1996; Kang et al., 2005) caused the same phenotype alterations as dnFGFR, showing a slight increase in the percentage of larvae totally deprived of pigment cells (Fig. 1.15 D,D',F). Conversely, the constitutive active form, Ets:VP16 (Cress and Triezenberg, 1991; Hall and Struhl, 2002), led to the formation of extranumery pigment cells (up to four) in the larval sensory vesicle (Fig. 1.15 E,E').
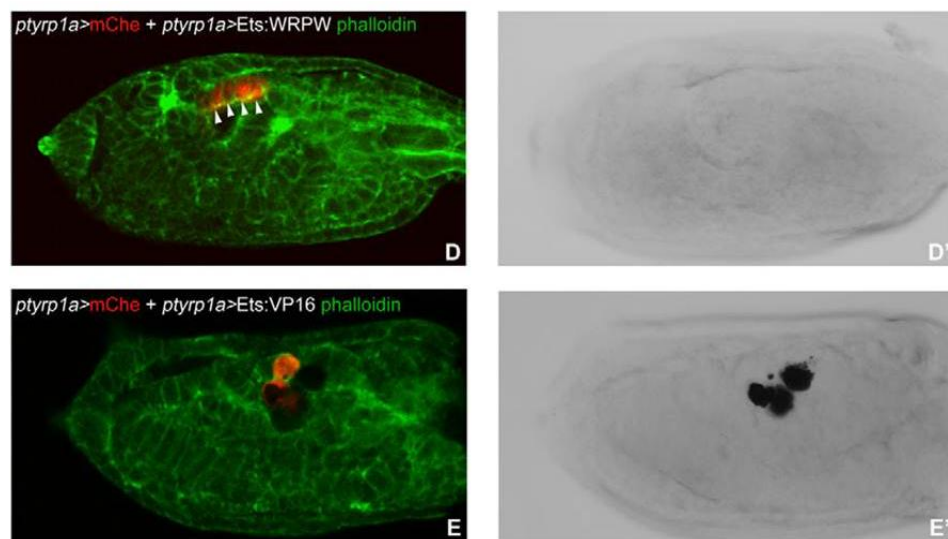
Fig. 1.15 *Ci-Ets1/2* is required for pigment cell precursors induction. Double transgenic ptyrp1a>mChe plus ptyrp1a>Ets:WRPW (D,D') and ptyrp1a>mChe plus ptyrp1a>Ets:VP16 (E,E') larvae observed with confocal and transmitted light, respectively. Pigment cells are absent (D'), there are four grouped cells in the sensory vesicle (D, arrowheads) and accessory pigment cells are present in the sensory vesicle (E'). Adapted from (Squarzoni P., et al., 2011).

1.5 Affymetrix GeneChip Microarrays

GeneChip arrays are the result of the combination of a number of technologies, design criteria, and quality control processes; the technology relies on standardized assays and reagents, instrumentation (fluidics system, hybridization oven and scanner), and data analysis tools that have been developed as a single platform. The key assay steps are outlined in Fig. 2. These instruments are designed to work together and, with the exception of the hybridization oven, are directed by the GeneChip operating software (GCOS). After hybridization the arrays are transferred to the fluidics station. The fluidics station performs washing and staining operations for GeneChip microarrays, a crucial step in the assay that impacts data consistency and reproducibility.

Unbound nucleic acid is washed away through a combination of low and high stringency washes. In the next step, bound target molecules are ''stained'' with a fluorescent streptavidin–phycoerythrin conjugate (SAPE), which binds to the biotins incorporated during target amplification. At the end of wash and stain steps, the arrays are ready for scanning. During the scan process a photomultiplier tube collects and converts fluorescence values into an electronic signal, which is then converted into the corresponding numerical values. These numerical values represent the fluorescence intensities, which are stored as pixel values that comprise the image data file (.dat file). The next step in analysis is the assignment of pixels that make up the image (.dat) file to the appropriate feature that are stored in a .CEL file.

Among microarrays, the single-channel Affymetrix GeneChip platform (http://www.affymetrix.com) is by far the most popular; in this technology each transcript is typically measured by a set of 11-20 pairs of 25 bases-long probes, collectively referred to as "probeset". For every "perfect match" probe (PM), Affymetrix chips contain a "mismatch" counterpart (MM), with a single nucleotide change in the middle of the PM probe sequence. The role of MM probes, located adjacent to the respective PM, is to measure probe-specific background signal associated to any perfect-match signal intensity. In general, the process of obtaining a single gene expression value out of raw probe intensity measurements is called "microarray preprocessing"; an expression summary value is calculated using a three-step procedure. (1) 'Background correction', which removes the unspecific background intensities of the scanner images; (2) 'normalization', which reduces the undesired non-biological differences between chips and normalizes the signal intensity of the arrays; (3) 'summarization', which combines the multiple probe intensities associated to the same gene to a single expression value (Fig 1.16).
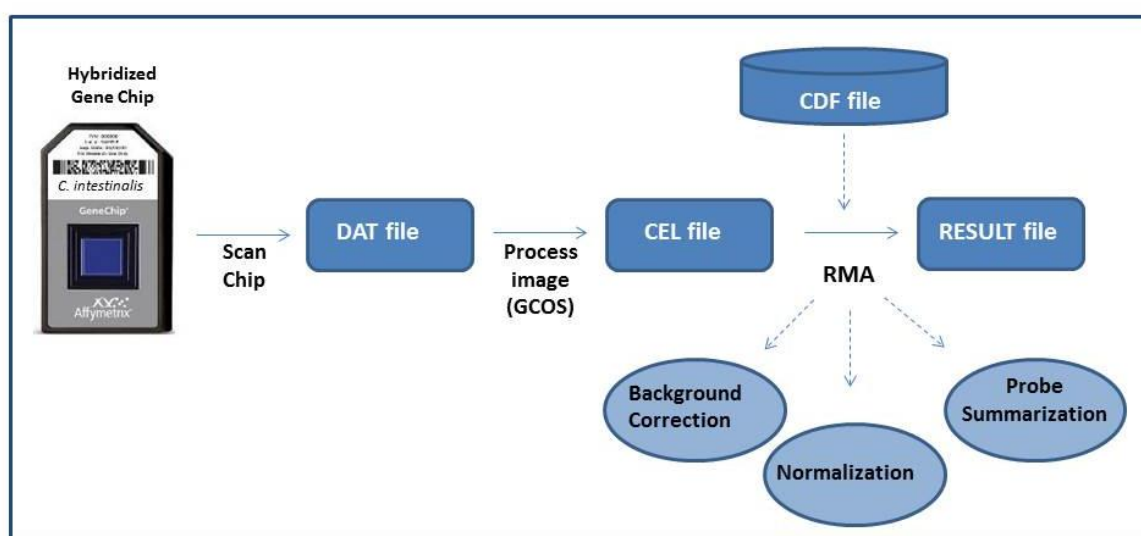
Fig.1.16 Affymetrix Data flow and Pre Processing Analysis.

Before any kind of microarray data can be analysed for differential expression several steps must be taken. Raw data must be quality assessed to ensure its integrity. Unprocessed raw data will always be subject to some form of technical variation and thus must be preprocessed to remove as many unwanted sources of variation as is possible, to ensure that results are of the highest attainable level of accuracy. Many different methods or combinations of methods for pre-processing were proposed over the years (Irizarry et al., 2006; Millenaar et al., 2006). The most popular manufacturer-provided method, MAS5 (Hubbell et al., 2002), uses a scale normalization approach, then corrects the background by subtracting the mean intensity of the lowest 2% spots in every microarray region, and then MM intensities from the respective PM ones. Wherever MM intensity is higher than a PM one, in order to avoid negative signal intensities typical of the MAS5 predecessor, MAS4 (Zang et al., 2008), MAS5 replaces the MM signal with an "idealized mismatch" value (IM) derived from other values in the same probeset. To extract final probeset intensities, MAS5 calculates a robust average (Tukey's biweight) of all the contained probes (Giorgi et al., 2010). Many alternative techniques have challenged MAS5 supremacy for preprocessing. Being a single-array technique, MAS5 doesn't model probes' behaviour across different samples, and therefore suffers from high variance and is theoretically less robust than multi-array algorithms (Wu and Irizarry, 2004). One of the most popular multi-array normalization techniques are RMA (Robust Multichip Average) (Irizarry et al., 2003), a method that doesn't use information contained in MM probes and calculates background signal by performing a modelled global correction of all PM intensities. Then it applies a quantile normalization step and a median polish summarization, which accounts for probe intensities over multiple arrays. Like most pre-processing algorithms, RMA performs 3 steps:

Background Correction

The RMA method begins by computing background correction of the intensity reading for each spot. The data recorded with the microarray technique are characterized by high levels of noise induced by the preparation, hybridization and measurement processes. Noise originates from chip fabrication tolerances, efficiency of RNA extraction and reverse transcription, background intensity fluctuations, non-uniform target labeling, temperature fluctuations, pipette errors, hybridization efficiency and scanning deviations. Also chemical effects may disturb the target signal in the data.

The background correction is a non-linear correction, done on a per-chip basis. It is based on the distribution of PM values amongst probes on an Affymetrix array. The background corrected intensities are computed in such a way that all background-corrected values must be positive.

Normalization

Subsequently the background-corrected intensities are normalized; the challenge of normalization is to remove as much of the technical variations as possible while leaving the biological variation untouched. There will always be slight discrepancy between the hybridisation processes for each array and these variations tend to lead to scaling differences between the overall fluorescence intensity levels of various arrays. The normalization used in RMA is quantile normalization method developed by (Bolstad et al., 2003).

Summarization

Once the probe-level PM values have been background-corrected and normalized, they need to be summarized into expression measures; in *Ciona intestinalis* GeneChip arrays 16 different PM are present to target 16 separate 25 base long sections of a target genes mRNA. The final step is to summarise the data from these 16 separate probes into an unique expression value for the gene in question. To achieve the summarization, the expression level is estimated using a robust procedure called Median Polish, an ad hoc procedure developed by Tukey (Tukey, 1977).

1.6 Aim of the Project

The aim of my thesis is to elucidate the transcriptional cascade responsible for *Ciona* pigment cell precursors (PCPs) formation by a microarray analysis. Recently, it has been demonstrated that the FGF signaling pathway, via the MAPK cascade, is the inductive signal responsible for conditional pigment cell precursor specification in *Ciona* (Squarzoni et al., 2011). Here, I focused the attention on extending the characterization of FGF cascade with the final aim to depict the gene regulatory network controlling Central Nervous System (CNS) differentiation and pigmented cell specification in *C. intestinalis*. To this aim, I employed a transcriptomic approach, using a microarray analysis, by exploiting a targeted interference of the FGF signalling together with the enrichment of cells of interest by Fluorescence Activated Cell Sorted (FACS). Target molecular interference of FGF signalling represent a powerful system for deciphering the downstream target gene network that could be directly or indirectly regulated by the cascade. FGF signalling can be either potentiated or disrupted leading to a strongly impaired pigment cell specification. The combination of cell sorting, cell-lineage specific transcription profiling and targeted molecular manipulations offers new technical challenges facilitating the identification of novel candidates specifically expressed in pigment cell lineage involved in FGF/MAPK/Ets regulatory architecture. Transcript profiling offers the largest coverage and a wide dynamic range of gene expression

information. Microarray data analysis allows the simultaneous measurement of the expression of many thousands of genes and is a valuable resource to produce lists of differentially expressed genes. Meanwhile the enormous amount of information generated by high throughput profiling technologies needs to be managed, integrated and interpreted with appropriate bioinformatic knowledge. A 'reverse-engineering' is a good approach to infer, from gene expression data, the regulatory interactions among genes using computational algorithms. New candidates co-expressed with melanogenic genes and involved in FGF signaling regulatory network could be computationally selected from a tissue-specific transcriptional network and neighbors analysis. Moreover clustering, such as hierarchical and k means, are unsupervised methods to group and classify samples and genes according to similar expression patterns; once identified cluster of genes with a similar expression values, there are several open source tools available to analyse them, such as oPOSSUM that permit to identify statistically the over-representation of conserved transcription factor binding sites in the promoter regions of a set of co-expressed genes. Functional enrichment such as metabolic pathways and gene ontology functional classifications integrates the normalized array data and their annotations. Bioinformatic pipelines for array data analysis and several tools represent a good opportunity for the efficient integration and interpretation of large datasets and to select unknown functional targets involved in the gene regulatory networks controlling *Ciona* CNS differentiation and PCPs formation.

CHAPTER 2

METHODS

2.1 Animal husbandry and embryos collection

Gravid C. intestinalis adults were collected in the Gulf of Naples, while for microarray experiments, that have been carried out at New York University in collaboration with Prof. Lionel Christiaen, *C. intestinalis* adults were collected at the Half Moon Bay marina (CA, USA). Adults were maintained under constant illumination in a sea water tank equipped with appropriate water circulation and filtering system. Ripe oocytes and sperm were collected surgically and kept separately until in vitro fertilization. Gametes collected from the several animals gonoducts, have been used for microarray experiments, electroporation or in situ hybridization experiments. Embryos were raised in Millipore-filtered seawater at 16-28° C. Only the batches in which at least 80% or more of the embryos developed normally were selected for the experiments.

2.2 Chemical dechorionation and in vitro fertilization

In order to eliminate chorion and follicular cells surrounding and protecting the eggs, a chemical dechorionation has been effectuated in a Petri dish with a thin layer of 1% agarose in MFSW on the bottom, putting the eggs for 5-6 minutes in a pH 10 solution of Thioglycolic acid (1%) and Proteinase E (0.05%) in MFSW. During this time, the eggs have been shaken continuously in this solution, using a glass pipette, to remove the chorion and the follicular cells surrounding the eggs. After this step the eggs have been washed several times to remove the residual solution and then fertilized in vitro with sperm collected from two or more individuals to avoid self-sterility problems. After 10 minutes, fertilized eggs have been washed 2-4 times to eliminate the exceeding sperm

2.3 Transgenesis *via* electroporation

The fertilized eggs have been immediately transferred in a solution containing 0.77 M Mannitol and 50-100 μg of DNA. The electroporation has been carried out in Bio-Rad Gene Pulser 0.4 cm cuvettes, using Bio-Rad Gene Pulser II with extension capability, at

constant 50 V and 500-800 μF, in order to have a time constant of 14-20 m/seconds. The embryos have been allowed to develop until the desired stage on 1% agarose coated Petri dishes, at 16-18°C. Depending on the electroporated constructs and on the presence of diverse reporter genes, the embryos have been treated in different ways.

2.4 GFP expressing embryos observation

The embryos and larvae have been observed in vivo, directly transferring them on slides for the observation at the microscope of fluorescence and phenotypic abnormalities. To avoid movement, tailbud and larva stages have been sedated, putting some menthol crystals in the Petri dishes, before the montage on the slides.

For direct GFP visualization, embryos were fixed for 20 to 30 minutes in MEM-FA, rinsed in PBS, 0.15% triton-X100, 0.05% tween-20 for 1 hour. All fluorescent samples were mounted in 50% glycerol /PBS supplemented with 2% DABCO (Sigma, #D2522). The embryos have been placed on a microscope slide. A cover slide (18x18mm) with some plasticine at the corners have been positioned on the top of the drop of embryos, and pressed until the volume resulted uniformly distributed. DIC and fluorescent images have been taken with a Zeiss Axio Imager M1 microscope equipped with an Axiocam digital camera. For confocal images, embryos have been analyzed with a Zeiss confocal laser scanning microscope LSM 510. Stacks of confocal images were processed using Image J for color merging, 3D- and Z projections.

2.5 DNA gel electrophoresis

Preparative and analytic DNA gel electrophoreses experiments have been performed on 1% or 1,5% of agarose gel in 1x TAE Buffer (Stock solution 50x TAE Buffer: 252 g of Tris base; 57.1 ml glacial acetic acid; 100 ml 0.5M EDTA; H2O to 1 liter), considering the length of the DNA to be run and adding 0.5 μg/ml Ethidium Bromide. Suitable DNA markers have been used to check DNA concentration (1x Lambda DNA/HindIII Marker 2, Fermentas) or length (1x GeneRuler™ 1Kb DNA Ladder; 1x GeneRuler™ 100 bp DNA Ladder, Fermentas). The DNA gel electrophoreses have been performed with constant voltages ranging from 60 to 100 V.

2.6 DNA gel extraction

Digested and PCR amplified fragments have been extracted from gel cutting them with a sterile sharpen blade, using the GenElute™ Gel Extraction Kit (Sigma-Aldrich), following

the manufacturer's instructions. After the extraction, the concentration has been estimated by gel electrophoresis.

2.7 DNA digestions with restriction endonucleases

Analytic and preparative plasmid DNA digestions have been performed. DNA of interest have been digested with the appropriate restriction endonucleases in total volumes of at least 20 times more than the enzyme volume used. The digestion reaction has been prepared as follows: the solution contained the desired amount of DNA, suitable Restriction Enzyme Buffer (1/10) (Roche; New England Biolabs; Amersham), Restriction Enzyme/s (5 units enzyme per 1 μg of DNA). Reaction specific temperatures have been used as suggested by manufacturer's instructions.

2.8 Digested plasmids' purification

To eliminate protein contaminations, the DNA has been purified with 1 volume of phenol:chloroform:isoamylic alcohol (25:24:1), vortexed vigorously and centrifuged at 13000 rpm for 5 minutes at 4°C. The soluble phase has been recovered and 1 volume of chloroform:isoamylic alcohol (24:1) has been added; the sample has been vortexed vigorously and centrifuged at 13000 rpm for 5 minutes, at 4°C. The aqueous phase has been recovered and the DNA has been precipitated adding 3 volumes of ethanol 100% and 1/10 volumes of Sodium acetate 3M pH 5.2. The sample has been mixed and stored over night at -20°C or 1 hour at -80°C. Then, it has been centrifuged at 13000 rpm for 15 minutes, at 4°C. The precipitated DNA has been washed with ethanol 70% (sterile or DEPC-treated), centrifuging at 13000 rpm for 15 minutes at 4°C. The ethanol has been removed and the sample has been air-dried at R.T. At the end, the DNA has been diluted in a suitable volume of H2O (sterile or DEPC-treated), and its concentration has been evaluated by gel electrophoresis, using 1X Lambda DNA/Hind III Marker 2 (Fermentas), and using a spectrophotometer (Nanodrop 1000, Thermo SCIENTIFIC), reading the values at the wavelengths of 230, 260 and 280 nm and calculating the ratio between 260/230 nm and 260/280 nm to ascertain respectively the absence of chemical (phenol, ethanol) and protein contamination.

2.9 DNA dephosphorylation

DNA dephosphorylation has been conducted as follows: a convenient amount of double stranded linearized DNA has been incubated with 1U of Calf Intestinal Alkaline Phosphatase enzyme (CIP) (Roche) per 1 pmol 5' ends of double stranded linearized DNA,

in 1X CIP dephosphorylation buffer (Roche), at 37°C for 30 minutes. After this time, a second aliquot of CIP as been added, and the reaction has been carried on for further 30 minutes, at 37°C. Subsequently, the dephosphorylated DNA has been purified by phenol: chloroform/chloroform extraction (see par. 2.8).

2.10 DNA Ligation

Each ligation reaction has been carried out in a final volume of 10 μl containing: 50-100 ng of linearized DNA vector; 3-5 fold vector moles of DNA insert;1X T4 DNA Ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$, 10 mM dithiothreitol, 1 mM ATP, pH 7.5) and 1 unit/μl of T4 DNA Ligase (New England Biolabs). Generally, the optimal proportion of 1 : 3, vector : insert, has been used in the reaction. The reaction mix has been incubated at 16 °C overnight or 1,5 hour at R.T., and used to transform competent bacteria.

2.11 Bacterial cells electroporation

By this approach it is possible to transform bacterial cells with plasmids containing DNA of interest. Briefly, the circular plasmid DNA and competent bacterial cells, prepared by the Molecular Biology Service of the Stazione Zoologica A. Dohrn in Naples, were placed in an electrocuvette. The electrocuvette was subjected to an electric pulse at constant 1.7 V using a Bio-Rad Gene Pulser™ electroporation apparatus. Then the cells were placed in 1 ml of SOC (tryptone 20 g/l, yeast extract 5 g/l, 5 M NaCl2 ml/l, 1 M KCl 2.5 ml/l, 1 M MgSO4 10 ml/l, 1 M MgCl2 10 ml/l, 1 M glucose 20 ml/l) shaking at 270 rpm at 37 °C for 1 hour, plated on LB solid medium (NaCl 10 g/l, bactotryptone 10 g/l, yeast extract 5 g/l, agar 15 g/l) in the presence of the specific antibiotic (50 μg/ml) to which the plasmid is resistant and grown at the same temperature overnight.

2.12 PCR Screening

It is possible to carry out a PCR reaction using as template the DNA of a single bacterial colony and in the same time grow the colony. Half of each single colony was placed in a PCR mixture described below, and half grown in 3 ml of LB liquid medium in the presence of the suitable antibiotic (50μg/ml), 8-12 hours shaking at 270 rpm, at 37 ºC.

The PCR reaction have been set in a total volume of 20 μl, with the following composition: 1x PCR buffer (Roche); 0.2 mM dNTP mix (dATP, dTTP, dCTP, dGTP); 1 pmol/μl of each forward and reverse suitable oligonucleotides; and 0.025 U/μl Taq DNA polymerase (Roche; Biogem). PCRs have been carried out with the following protocol:

1. First step (one cycle): 5 minutes at 95°C

2. Second step (repeated for 25 cycles):

DNA denaturation; 1 minute at 95 ºC

 Oligonucleotides annealing: 1 minute 52-60°C (the temperature used in this step has been set at least 2-3°C below the melting temperature of the oligonucleotides).

Polymerization: 72°C for a suitable time (the polymerization time has been calculated considering the desired amplified fragment length and the Taq DNA Polymerase processivity, that is around 1 Kb/minute).

By gel electrophoresis analysis, the samples presenting a band of expected size have been identified and plasmid DNA have been purified from the corresponding bacterial colonies.

2.13 DNA Mini- and Maxi-preparation

A single bacterial colony containing the plasmid DNA of interest was grown in a suitable volume of LB in the presence of the appropriate antibiotic shaking at 37°C overnight. The Sigma-Aldrich Plasmid Purification kit, based on alkaline lyses method, was used to isolate the plasmid DNA from the cells according to the manufacture's instruction.

2.14 Preparation of constructs used for microarray experiments
pTyrp1/2a>dnFGFR

In order to obtain pTyrp1/2a>dnFGFR construct, I used the construct MESP>FGFRDN kindly provided by Prof. Lionel Christiaen, New York University, containing the coding sequence for the dominant negative FGF receptor. MESP>FGFRDN has been double digested with XbaI and NotI restriction enzymes (Roche), to eliminate the MESP promoter sequence. The digested plasmid, containing the sole dnFGFR, has been analyzed on 1% agarose and then purified by gel extraction (GenElute™ Gel Extraction Kit, Sigma-Aldrich). The promoter fragment for pTyrp1/2a has been obtained from double digests using XbaI and NotI, of pTyr>dnFGFRmCherry (already available in the lab); the two XbaI-NotI digested fragments have been ligated (T4 DNA Ligase, New England Biolabs), using an optimal proportion of 1 : 3, vector : insert. gel and then ligated using a 1:4 ratio vector/insert. The recombinant plasmids were transformed by electroporation, as already described (Par. 2.11), in Top10 bacterial competent cells. The bacterial cells were plated and grown at 37°C overnight.

pTyrp1/2a>EtsVp16

pTyrp1/2a>dnFGFR has been double digested with NotI and EcoRI restriction enzymes (Roche Applied Biosystems), to eliminate dnFGFR. The digested plasmid, containing pTyrp1/2a, has been isolated by gel extraction (GenElute™ Gel Extraction Kit, Sigma-

Aldrich). MESP>EtsVp16 and (generous gift from Prof. Lionel Christiaen, New York University, New York) have been digested with NotI and EcoRI restriction enzymes (Roche Applied Biosystems) and the fragment containing EtsVp16 has been purified by gel extraction. The NotI - EcoRI EtsVp16 digested fragment has been ligated into the NotI - EcoRI plasmid pTyrp1/2a using T4 DNA Ligase with the optimal proportion of 1 : 3, vector :insert.

pTyrp1/2a>GFP

pTyrp1/2a>dnFGFR has been double digested with NotI and EcoRI restriction enzymes (Roche Applied Biosystems), to eliminate dnFGFR. The digested plasmid, containing pTyrp1/2a, has been isolated by gel extraction. The vector pSP72 Hβg GFP, with the Human β-globin minimal promoter included upstream the coding sequence for GFP to increase the efficiency of the proteins expression, has been linearized by HindIII digestion, and dephosphorylated by Calf Intestinal Phosphatase (CIP), as previously described in Par. 2.8. The HindIII digested pTyrp1/2a fragment has been subsequently ligated with HindIII-linearized pSP72 Hβg GFP, using an optimal proportion of 1 : 3, vector : insert. Finally, after the propagation of the construct by bacterial transformation, the obtained positive clones have been sequenced with oligo Sp6 to ascertain the 5'-3'orientation of the pTyrp1/2a fragment in the pSP72 pTyrp1/2a Hβg GFP construct.

pTyrp1/2a>2xGFP

To obtain the pTyrp1/2a>2xGFP construct, I double digested the plasmid pTyrp1/2a>dnFGFR with NotI and EcoRI to eliminate the coding sequence for dnFGFRDN and purified the band with pTyrp1/2a fragment by gel extraction. I inserted this fragment in the plasmid COE>2xGFP kindly provided by Prof. Lionel Christiaen, previously digested with NotI and EcoRI and purified by gel exctraction. The ligation reaction has been conducted using an optimal proportion of 1 : 3, vector : insert.

pMyoD>YFP

pMyoD>YFP transgene has been kindly provided by Prof. Lionel Chrsitiaen and was generated as follows: the 0.9kb MRF/MyoD upstream cis-regulatory DNA was PCR amplified from genomic DNA using

-MyoDF905 (5'-gcgctctagaggcttacgcatctcgagcgaac-3')

-MyoDNatB (5'-tttgcggccgcctctagagagatacacgtcatcg-3') primers

and cloned into the pCES vector containing lacZ as an XbaI-NotI fragment. The lacZ coding sequence was then replace by a Venus YFP NotI-EcoRI fragment amplified from pSD1:venus, using Not-venus-F2 (tttgcggccgccaacatggtgagcaagggcgaggag) and Venus-R2 (gctggaattcttacttgtacagctc).

pTyrp1/2a>LacZ

pTyrp1/2a>LacZ has been obtained using the same pTyrp1/2a fragment, double digest with NotI and EcoRI and purified by gel extraction, as described before. This fragment has been inserted in the pBlueScript 1230 vector, containing the coding sequence for β-Galactosidase, SV40 polyadenylation sequence and the human β-globin basal promoter, previously digested with NotI and EcoRI. The ligation reaction has been conducted using an optimal proportion of 1 : 3, vector : insert.

2.15 Embryo cellular dissociation for Fluorescence Activated   Cell Sorting

Embryonic cell suspensions for fluorescence activated cell sorting (FACS) were obtained as follows: stage-selected embryos were rinsed three times in artificial calcium and magnesium free sea water (CMF-ASW: 449mM NaCl, 33mM Na2SO4, 9mM KCl, 2.15mM NaHCO3, 10Mm Tris-Cl pH8.2, 2.5mM EGTA) at room temperature. Embryos were dissociated by thorough pipetting for 2 to 3 minutes in 1ml CMF-ASW supplemented with 0.1% trypsine (w/v, Sigma, T-4799). Trypsine activity was blocked by adding ice-cold CMF-ASW supplemented with 0.1% Bovine Serum Albumine (BSA, Sigma, A-3311). Cell suspensions were then rinsed three times with ice cold CMF-ASW+BSA (cell were pelleted at 4°C, by centrifugation at 3000rpm (~500-700g), for 2 minutes). Cell suspensions were maintained on ice and filtered into 5ml polystyrene round bottom tubes equipped with a 40μm cell-strainer cap (Falcon, 352235) prior to cell sorting.

2.16 Fluorescence Activated Cell Sorting

Samples were analyzed and sorted in a Coulter EPICS Elite ESP sorter (Coulter,Inc.,Miami FL). A JDS Uniphase, Inc.(New York University, NY), 15mW air cooled laser tuned for 488nm light emission was used for light scattering and excitation of GFP and YFP. Individual cell sorting was achieved using a 100um flow cell nozzle equipped with a confocal lens assembly for enhanced amplification of cell fluorescence GFP and YFP fluorescence distributions were gated on the population of cells. Special attention was devoted to exclude debris and cell aggregates by gating. GFP fluorescence was collected using a 500nm Dichroic long pass filter and a 510nm discriminating optical filter (Omega Optical, Brattleboro, Vermont). YFP fluorescence was collected using 550nm Dichroic long pass filter and a 550nm discriminating optical filter (Omega Optical, Brattleboro, Vermont). For every sort, a non-fluorescent cell suspension as well as GFP only and YFP only samples were analyzed to determine the optimal settings for sorting. Spectral overlap was adjusted using these samples to compensate for the GFP fluorescence

component detected by the YFP detector and for the YFP fluorescence component detected by the GFP detector (Fig. 2.1 and Fig 2.2). Collimated GFP and YFP fluorescence was separated using 500nm dichroic long pass filter. GFP fluorescence (fluorescence lower than 500nm) was reflected towards a photomultiplier detector (PMT) with a 510nm discriminating filter in front of it. YFP fluorescence was detected from the transmitted 500nm dichroic long pass filter using a 550nm dichroic long pass optical filter. A 550 nm discriminating optical filter was placed in the light path towards the esignated YFP fluorescence PMT detector.

2.17 RNA extraction and RNA quality detection

Total RNA extraction was performed immediately after the cell sorting using the RNAqueous-micro kit (Ambion, #1931). Frozen samples were thawed at room temperature, 500µl CMF-ASW was added and cells were pelleted by centrifugation for 4 minutes, 4000rpm, at 4°C. The supernatant was quickly removed and replaced with 200µl lysis buffer. Cells were lysed by thorough vortexing. RNA extraction was performed as per manufacturer recommendations and eluted in 2x6.5µl of pre-warmed (70°C) elution buffer. Total RNA (1µl) were analyzed on the BioAnalyzer machine using the Eukaryote total RNA Pico Series II kit using the mRNA Smear Nano program following the manufacturer's instructions following manufacturer's recommendations (Agilent Technologies, #5067-1513). The integrity of total RNA is estimated by RNA Integrity Number (RIN) values, calculated by an algorithm that assigns a 1 to 10 RIN score, where level 10 RNA is completely intact values; in this way integrity is no longer determined by the ratio of the ribosomal bands, but by the entire electrophoretic trace of the RNA sample. This includes the presence or absence of degradation products (Fig. 3.4 A,B).

2.18 RNA amplification and cDNA preparation for Microarray experiments

RNA amplification and probe synthesis were performed using The Ovation Pico WTA System following manufacturer's recommendations (NuGen technologies); this kit is ideal for amplification of compromised RNA samples, because powered by Ribo-SPIA technology, a rapid and sensitive RNA amplification process that allows to prepare microgram quantities of cDNA starting from 500 pg to 50 ng total RNA. In particular, First Strand cDNA generation, using primers have a DNA portion that hybridizes either to the 5´ portion of the poly(A) sequence or randomly across the transcript. Reverse Transcriptase (RT) extends the 3´ DNA end of each primer generating first strand cDNA. The resulting cDNA/mRNA hybrid molecule contains a unique RNA sequence at the 5´ end of the

cDNA strand. Fragmentation of the mRNA within the cDNA/mRNA complex creates priming sites for DNA polymerase to synthesize a second strand, which includes DNA complementary to the 5´ unique sequence from the first strand chimeric primers. The result is a double-stranded cDNA with a unique DNA/RNA heteroduplex at one end. SPIA amplification is a linear isothermal DNA amplification process that uses a SPIA DNA/RNA chimeric primer, DNA polymerase and RNase H in a homogeneous isothermal assay. RNase H is used to degrade RNA in the DNA/RNA heteroduplex at the 5´ end of the first cDNA strand. This results in the exposure of a DNA sequence that is available for binding a second SPIA DNA/RNA chimeric primer. DNA polymerase then initiates replication at the 3´ end of the primer, displacing the existing forward strand. The RNA portion at the 5´ end of the newly synthesized strand is again removed by RNase H, exposing part of the unique priming site for initiation of the next round of cDNA synthesis. SPIA amplification final result is a rapid accumulation of cDNA with sequence complementary to the original mRNA. An average mRNA amplification of 15,000-fold is observed with 500 pg starting total RNA.

The fragmented and labeled single-stranded cDNA target has been generated using the Encore Biotin Module, from NuGen techlonogies, following manufacturer's recommendations (#4200-12). Briely it produces fragmented and labeled single-stranded cDNA ranging from 50 to 100 bases ready for hybridization to GeneChip Affymetrix arrays.

2.19 Target Preparation for Affymetrix GeneChip Eukaryotic Array Analysis

Amplified, fragmented and biotin-labeled cDNA targets generated using the Encore Biotin Module are prepared for analysis on GeneChip arrays according to the Affymetrix GeneChip Expression Analysis Technical Manual (P/N 702232 Rev. 2). To prepare target for a single standard array, mix at room temperature the amount of target cDNA (25μL) and volumes of hybridization cocktail components indicated in the Table 1 below. Heat denaturation for the hybridization cocktail have been done at 99 °C for 2 minutes, then 45 °C in a heat block for 5 minutes as suggested by the Affymetrix standard protocol. The probe array filled with Pre-Hybridization Mix have used for hybridization to the Ciona intestinalis custom-design GeneChip for 18 hours using standard Affymetrix procedures at the New York Univesity, NYU's Center for Genomics & Systems Biology. The Fludics protocol selected was for ST arrays (FS450_0007 Gene arrays).

Table 1. Hybridization, Cocktail Assembly and Fluidics Protocols for Single GeneChip Arrays using Affymetrix kit (Affymetrix P/N 900720)		
Component		**Final Concentration**
Fragmented, biotin-labeled amplified cDNA	25 µL	
Control oligonucleotide B2 (3nM)	1.8 µL	50 pM
20x Eukatiotic hybridization controls (bioB, bioC, bioD, cre)	5.5 µL	1.5, 5, 25 and 100 pM Respectively
2x Hybridization buffer	55 µL	1x
100% DMSO	11 µL	10%
Water	11.6 µL	N/A
Final Volume	110 µL	

2.20 Affymetrix GeneChip Microarrays platform

Affyemtrix GeneChip arrays are the result of the combination of a number of technologies, design criteria, and quality control processes. In addition to the arrays, the technology relies on standardized assays and reagents, instrumentation (fluidics system, hybridization oven and scanner), and data analysis tools that have been developed as a single platform. The key assay steps are outlined in Fig. 2.1.
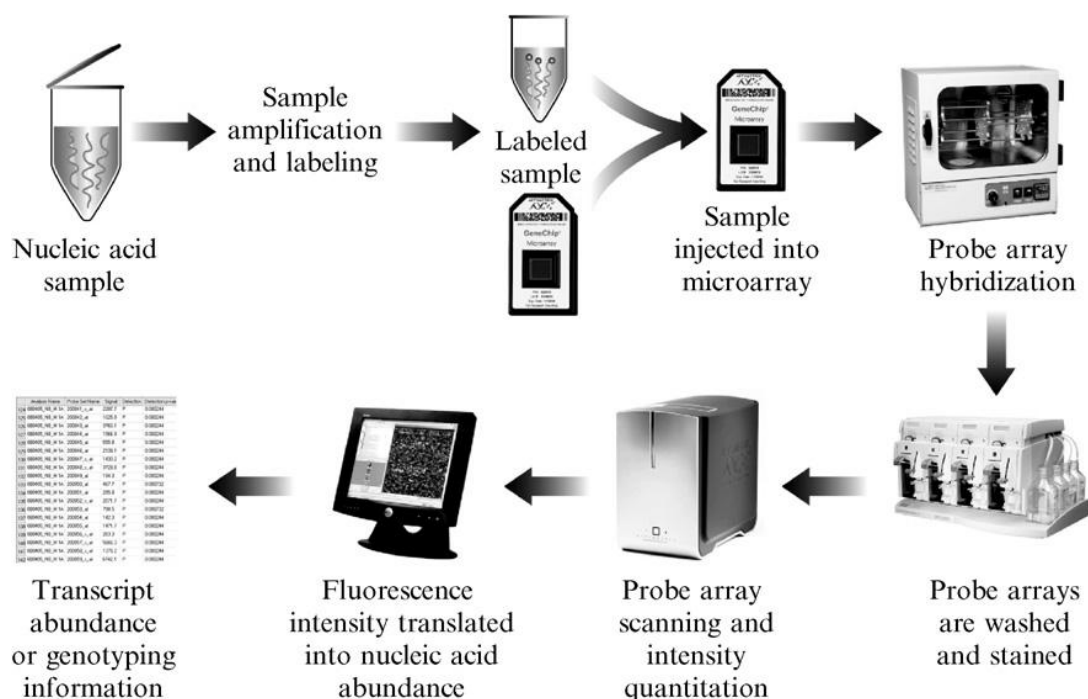
Fig. 2.1 Flowchart of a GeneChip System microarray experiment: adapted from Dalma-Weiszhausz D.D., et al.,2006.

2.21 Microarray data analysis

Microarray data analysis has been done using GeneSpring GX 12.1, a popular gene expression analysis software that provides powerful, accessible statistical tools for fast visualization and analysis of expression and genomic structural variation data. The first operation in GeneSpring software to analyze microarray data is to create a technology that is defined as the species specific genome printed on microarray chip including

biological and other information about the probes, eg., Entrez gene ID, GO accession, Gene Symbol etc.

Raw expression values (.CEL files) for each probe and each biological replicate (four for each experimental condition) were used for normalization and computation of probe set expression estimates using the robust multi-chips analysis (RMA) algorithm (Irizarry et al., 2003). Quality Control analysis have been performed on Normalized data, comprising Box whisker plots for all normalized data, Scatter plots among biological replicates, 3D-Principal Component Analysis (PCA), Correlation analysis. For individual probe sets, p-values for expression changes between selected comparisons were computed using a statistical analysis on the normalized data, t-test and one-way ANOVA; Fold Change analysis have been performed on probe sets showing a p-values below 0.05.

2.22 Transcriptional regulatory Networks

Transcriptional regulatory networks are used to represent pairs of transcripts that coherently change their expression levels across a set of different conditions (i.e. co-expressed genes). Transcriptional regulatory networks connect transcripts that exhibit high statistical dependence by observing changes in their responses across all the experiments in the dataset. As a measure of similarity the pair-wise correlation (Pearson or Spearman) is often used.

The Spearman Correlation Coefficient (SCC) is a nonparametric measure of statistical dependence between two variables. It measures the strength of a monotonic relationship between paired data. Usually, SCC is denoted by the Greek letter ρ or r_S.

The Spearman correlation coefficient is defined as the same of Pearson Correlation Coefficient (PCC) but between the ranks of the variables. Let us two variables X and Y of size n, we first have to convert the n raw scores X_i, Y_i in ranks, x_i, y_i and then we can compute r_S as:

$$r_S = \frac{\Sigma_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_i(x_i - \bar{x})^2 \, \Sigma_i(y_i - \bar{y})^2}}$$

If there are identical values (i.e. rank ties or value duplicates) for these is assigned a rank equal to the average of their positions in the ascending order of the values.

The significance of SCC can be tested using the following formula (Kendall et al., 1994)

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

which is distributed approximately as Student's t distribution with $n-2$ degrees of freedom under the null hypothesis.

2.23 PCR amplification from DNA

The PCR reaction has been performed in a total volume of 50 μl, using about 100 ng of genomic DNA as template, 0.2 mM of dNTP mix (dATP, dTTP, dCTP, dGTP), 1x PCR Buffer (Roche), 0.05 U/μl of Taq DNA Polymerase (Roche) and 2 pmol/μl of each forward and reverse suitable oligonucleotides. The PCR amplification reaction from plasmid DNA has been performed to amplify fragments of interest from plasmid DNA as template. The total volumes ranged from 50 μl to 100 μl, depending on the needs. The PCR amplification program has been set as follows:

1. First step (one cycle): 5 minutes at 95°C for genomic DNA and 1 minutes at 95°C for plasmid DNA

2. Second step (repeated for 35 cycles):

DNA denaturation: 1 minute at 95°C.

Oligonucleotides annealing: 2 minutes 50°C for genomic DNA and 1 minute 52-54°C for plasmid DNA (the temperature used in this step has been set at least 8-10°C below the melting temperature of the oligonucleotides).

Polymerization: 72°C for a suitable time (the polymerization time has been calculated considering the desired amplified fragment length and the Taq DNA Polymerase processivity, that is around 1 Kb/minute).

Final elongation step: 10 minutes at 72°C for genomic DNA and 5 minutes at 72°C for plasmid DNA. The amplified fragments have been separated from the genomic DNA and from the dNTPs and oligonucleotide excess, by gel electrophoresis, using, as fragment length marker, 1x GeneRuler™ 1Kb DNA Ladder (Fermentas). The fragments have been isolated and purified by gel extraction (QIAquick Gel Extraction Kit, QIAGEN; GenElute™ Gel Extraction Kit, Sigma). The concentration has been evaluated by gel electrophoresis, using, as concentration marker, 1x Lambda DNA/HindIII Marker,2 (Fermentas).

2.24 Oligonucleotides' synthesis

All used synthetic oligonucleotides were prepared with a device Beckman SM-DNA Synthesizer by the Department of Molecular Biology of the Stazione Zoologica Anton Dohrn in Naples.

2.25 DNA Sequencing

The DNA sequences have been obtained using the Automated Capillary Electrophoresis Sequencer 3730 DNA Analyzer (Applied Biosystems, Foster City, CA) by the Molecular Biology Service of the Stazione Zoologica Anton Dohrn in Naples.

Oligonucleotide sequences used are:

- M13 Reverse 5'-CAGGAAACAGCTATGAC-3'

- SP_6 forward 5'-ATTTAGGTGACACTATAGAA-3'

- T3 forward 5'-ATTAACCCTCACTAAAG-3'

2.26 Ribonucleic probes' preparation

RNA labelling

The plasmid, containing the template to be transcribed, has been conveniently digested and purified, as described in pars. 2.14 and 2.15. 1 μg of purified, linearized DNA has been used as template for the ribonucleic probe synthesis. At this, transcription buffer (0.1%)

(Roche), Digoxigenin or Fluorescein labelling mix, containing 1 mM of ATP, CTP and GTP, 0.65 mM UTP and 0.35 mM DIG-11-UTP or 0.35 mM fluorescein-12-UTP (Roche), Sp6 or T7 RNA polymerase (2U/μl) (Roche), and Protector RNase inhibitor (1U/ μl) have been added. The reaction has been performed in a total volume of 20 μl (in H2O DEPC-treated). The mix has been briefly centrifuged and incubated for 2 hours at 37°C. Then, DNaseI RNase free (0.9U/μl) has been added and the sample has been incubated for 20 minutes at 37°C. Finally, the reaction has been stopped adding EDTA pH 8.0 (16 mM). The synthesized ribonucleic probes have been purified using the mini Quick Spin RNA Columns G-50 Sephadex (Roche), following manufacturer instructions. The ribonucleic probe quality has been checked by gel electrophoresis and the concentration quantification has been evaluated by Dot Blot analysis (see the next paragraph). The recovered samples have been immediately stored at -80°C till the use.

Ribonucleic probes' quantification by Dot Blot analysis

The concentration evaluation of the DIG-11-UTP or fluorescein-12-UTP incorporated in the ribonucleic probes has been estimated making serial dilutions of the sample ribonucleic probes and of a Control RNA of reference (Roche), in a buffer containing DEPC-treated H_2O:20x SSC:formaldehyde (5:3:2). 1 μl of each dilution has been blotted on a membrane Hybond-N (Amersham) and air-dried at RT. The RNA has been UV-crosslinked on the membrane with a Stratalinker for 30 seconds. The membrane, with the UV-crosslinked RNA on it, has been washed in blocking solution (5% BSA in 0.1 M maleic acid pH 7.5), for 30 minutes, shaking at RT. Subsequently, the membrane has been incubated with anti-DIG or anti-Fluo phosphate alkaline antibody (0.15 U/ml) (Roche) in blocking solution for 1 hour, shaking at RT. To remove unbound antibodies, the membrane has been washed twice in a solution containing 0.1 M maleic acid pH 7.5 and 0.15 M NaCl for 15 minutes, at RT. The membrane has been equilibrated in the detection solution (100 mM NaCl; 100 mM Tris pH 9.5; 50 mM $MgCl_2$, in H2O) for 5 minutes, at RT, and then incubated in the dark in the same solution in which Nitro-Blue Tetrazolium Chloride (NBT) (35 μg/ml) and 5-Bromo-4-Chloro-3'-Indolylphosphate p-Toluidine (BCIP) (175 μg/ml) have been added. The reaction has been monitored every 4-5 minutes and blocked at the suitable moment, putting the membrane under running water. The membrane has been dried on 3MM paper and the concentration of the DIG-11-UTP or fluorescein-12-UTP incorporated in the ribonucleic probes has been calculated from the comparison with the control RNA spots.

2.27 Whole Mount In Situ Hybridization (WMISH) assays

In situ hybridization was carried out on *C. intestinalis* whole mount embryos at different developmental stages fixed at RT 90 minutes or at 4°C over night, in a mixture containing: 4% paraformaldehyde, 0.1 M MOPS pH 7.5, 0.5 M NaCl.

The samples were washed 3x15 minutes in 1 ml PBT (PBS + 0.1% tween 20) at RT and incubated 30 minutes at 37°C in 1 ml PBT containing 2 µg/ml protease K for dechorionated embryos or 4 µg/ml for non dechorionated embryos, to increase the permeability to the cells and accessibility to mRNA target. The reaction was stopped by a wash in 2mg/ml glycine in PBT. After digestion, samples were post-fixed 1 hour at RT in 4% PFA+0.05% tween-20 in PBS 1x and then washed 3x15 minutes in PBT. Embryos were placed 10 minutes at same temperature in the pre-hybridization solution (50% formamide, 6xSSC, 0.05% tween 20) and finally 2 hour at 55°C in hybridization solution (50% formamide, 1x Denhardt's solution, 6xSSC, 0.05% tween 20, 100 µg/ml Yeast tRNA, 0,005% Heparine). Riboprobe (or Riboprobes for the double in situ hybridization) was added up to a final concentration of 0.5 ng/µl and the hybridization occurred overnight at the same temperature.

The following day a series of washes were carried out by varying the temperature and salinity conditions; embryos were, in fact, washed at 55°C in the following solutions: two times for 20 minutes in WB1 (50% formamide, 5x SSC, 0,1% SDS), two times for 20 minutes in WB1:WB2 and two times for 20 minutes in WB2 (50% formamide, 2x SSC, 0,1% Tween 20).

Subsequently they were treated with a 1 ml Solution A (10mM Tris-Cl, pH8.0, 0.5M NaCl, 5mM EDTA, 0.1% tween), two times for 5 minutes at RT. To remove aspecific RNA, not bound to the corresponding endogenous mRNA, the embryos were treated with RNase A [20 µg / ml] for 20 minutes at 37 °C in solution A, then washed one time in WB3 (2x SSC, 0.1% Tween20) for 5 minutes at RT and 2 times in WB3 at a temperature of 55°C.

Following they were incubated three times for 5 minutes in TNT (0.1M Tris, pH7.5, 150mM NaCl, 0.1% tween) at RT. At this point the detection of the probes has been carried out.

For underline single in situ hybridizations embryos were incubated in Blocking TNB buffer (100mM Tris pH7, 150mM NaCl, 1% Blocking Reagent, 0.2% Triton-100x) for 2 hours at R.T. At this point embryos were incubated, all night at 4°C, with the antibody anti-DIG in the ratio 1:2000 in Blocking TNB Buffer. The next day, the samples were washed at room temperature in TNT with the following modalities: one time for 5 minutes, four times for 20 minutes, one time for 40 minutes and three times for 10 minutes in TMN (100mM NaCl, 50mM MgCl2, 100mM Tris-Cl, pH9.5, 0.1% Tween20).

To identify the localization of the RNA of interest, labeled with DIG and recognized by anti-DIG alkaline phosphatase conjugated, are provided the appropriate substrates that will be converted by the phosphatase in a precipitate of blue.

The embryos are incubated, therefore, in 1 ml of TMN containing 4.5 µl of NBT (nitroblue tetrazolium) and 3.5 µl of BCIP (5-bromo-4-chloro-3-indolyl-phosphate). The time of formation of the precipitate are conditioned by the amount of antibody bound and, therefore, indirectly on the type of probe used. For this reason, at intervals of 30 minutes a few embryos are taken and observed, after being placed on a microscope slide, with a phase contrast microscope. When some signal is shown the color reaction is stopped using 1x PBT.

For <u>double in situ hybridization</u> embryos were first incubated in TNB blocking buffer for 2 hours at RT. At this point samples were incubated over night at 4°C, with anti-DIG Fab Fragments POD HRP diluted 1:400 in Blocking Buffer TNB.

The day after the following series of washes were carried out: four times for 15 minutes and two times for 5 minutes in TNT at RT.

To identify the localization of the RNA of interest marked with Digossiggenin and recognized by anti-DIG conjugated to alkaline peroxidase (HRP) a substrate is converted by HRP in fluorescent product. The embryos were incubated, therefore, in the 1x Plus Amplification Diluent (Perkin-Elmer) for 15 minutes at RT and subsequently 1:400 Cy3 diluted in the same solution for 1.5 hours at RT. After the reaction, embryos were washed nine times for 5 minutes in TNT at RT. In order to stop the first antibody reaction an incubation in 50% formammide, 2x SSC, 0,1% tween-20 for 10 minutes at 55°C was performed. The next step consisted in a series of three washes of 10 minutes in TNT and incubates, overnight at 4°C, with anti-Fluorescein HRP diluted 1:400 in Blocking Buffer TNB. The next day, the embryos were washed two times for 5 minutes in TNT at RT and then incubated in the 1x Plus Amplification Diluent (Perkin-Elmer) for 15 minutes at RT and subsequently treated with 1:400 Cy5 diluted in the same solution for 1.5 hours at RT. After this time the reaction was blocked by carrying out the following washings: 1 time for 5 minutes and 7 times for 10 minutes in TNT at RT. When desired embryos were treated three times for 5 minutes and then over night with 1:10000 DAPI in TNT, in order to highlight the nuclei of the cells.

The ribonucleic probes listed (Appendix II-Table 2) have been selected from microarray dataset for the synthesis of antisense ribonucleic probes starting from cDNA clones present in N. Satoh C. intestinalis gene collection 1 and Gateway library, available in the laboratory.

2.28 Preparation of constructs used for *Ci-Rab38/32/rab-rp1/ltd* promoter analysis

All the fragments selected for dissection analysis on *Ci-Rab38/32/rab-rp1/ltd* promoter were obtained by PCR amplification on genomic DNA using specific primers designed according to the sequence of *C. intestinalis* genome (http://genome.jgi-psf.org/ciona4/ciona4.home.html) (Table 3) and cloned in pCR®II vector (TOPO® TA Cloning Dual Promoter Kit, Invitrogen), following the manufacturer's indications. The pCR®II vectors, containing the *pCi-Rab38/32/rab-rp1/ltd* fragments (for simplicity called pRab38), have been sequenced using the oligo M13Rev and double digested using two combination of restriction enzymes based on the PCR product orientated insertion in pCR®II vector: XhoI and Hind III or HindIII and XbaI (Roche); these fragments have been purified by gel extraction (GenElute™ Gel Extraction Kit, Sigma-Aldrich) and subsequently cloned in pSP72 Hβg GFP, double digested with the same restriction endonucleases. The ligation reaction has been conducted by ligation (T4 DNA Ligase, New England Biolabs), using an optimal proportion of 1 : 3, vector : insert.

Table 3: Oligonucleotides used for the pfgf8/17/18 fragments' amplification from genomic DNA	
pRab38 oligonucleotide's name	*pRab38* oligonucleotide's sequence
pRab38A forward	5'-GTTGTCGCAACAAAAACTCC-3'
pRab38A reverse	5'-TGTTGCGATTATTCCACAAA-3'
pRab38B forward	5'-TAAGCGATGTTTCGCTTGTG-3'
pRab38B reverse	5'-AAACTGCCCCGACCCGGGAT-3'
pRab38C forward	5'-ATGCAACAAACTCCCTTTATT-3'
pPRab38C reverse	5'-CTCTTTTCGTCGTCGTCTTTA-3'
pRab38 D reverse	5'-CGCAATAATTATCCATCTGCTG-3'
pRab38 E forward	5'-GGCTTGGTTCGAAACTTCAG-3'

2.29 Isolation of the *Ci-Rab38/32/rab-rp1/ltd* cDNA

The full-length cDNA of *Ci-Rab38/32/rab-rp1/ltd* was found in Aniseed database (http://www.aniseed.cnrs.fr/) (Tassy et al., 2010) as cieg091j08. The correspondent cDNA clone was present in N. Satoh *C. intestinalis* gene collection 1, available in the laboratory, in plate 40 at position p15 (R1CiGC40p15). I picked the bacteria containing the plasmid with *Ci-Rab38/32/rab-rp1/ltd* full-length cDNA and let them grown in a suitable volume of LB in the presence of the appropriate antibiotic (1:1000), shaking at 37 ℃ for 16 hours. The plasmid DNA has been purified using the Sigma-Aldrich Plasmid Purification kit. In

order to verify that the clone has the right full-length cDNA, I sequenced the DNA using T3 oligonucleotide.

2.30 Preparation of the construct pTyr>Rab38/32

The full-length cDNA present in N. Satoh *C. intestinalis* gene collection are inserted in the vector pBlueScript KS II plus (Agilent Technologies). In order to create pTyr1/2a>Rab38/32, I amplified by PCR the *Ci-Rab38/32/rab-rp1/ltd* full-length cDNA (for simplicity called Rab38/32) using as template a plasmid containing the previously described clone from the library. As primers for the PCR reaction two oligonucleotides of about 30 bases in length, complementary to regions of interest and containing desirable restriction sites, in particular NotI at the 5'end and EcoRI at the 3'end, were used (Table 4). PCR reaction was performed as previously described. The amplified fragment was analyzed and extracted from 1% agarose gel as already described. The fragment was digested with EcoRI and NotI restriction enzymes, analyzed on 1% agarose gel and then ligated, using a 1:5 ratio, in the vectors pTyr>mCherry, previously digested with EcoRI-NotI. The reaction mixtures were incubated at 16 ° C overnight. The day after the T4 DNA ligase was inactivated at 65° C for 10 minutes and microdialized against sterile water for an hour, to remove the salts that could create problems in the subsequent phases. The obtained circular plasmids were transformed by electroporation in TOP10 electro-competent bacterial cells, prepared by the Molecular Biology Service of the Stazione Zoologica A. Dohrn of Naples (Dower, 1990) with an average conversion efficiency of more than $5\times10^{8}/\mu g$ of DNA. The bacterial cells were plated on solid medium containing ampicillin and incubated at 37°C for 18 hours. Some colonies were incubated in a suitable volume of LB in the presence of the appropriate antibiotic (1:1000), shaking at 37 ºC for 16 hours. The day after minipreps were performed and clones-sequences were analyzed by sequencing. The plasmid DNA has been purified using the Sigma-Aldrich Plasmid Purification kit. The DNA has been sequenced using the pTYRint Forward (5' TCACACATTTGTCTAGCACTTCCAAAG 3'), an oligonucleotide designed specifically at the end of *Ci-Tyrosinase* promoter fragment to ascertain the positively ligation with the full length cDNA.

Table 4: Oligonucleotides used to amplify Rab38/32 full-length cDNA	
Oligo's name	**Oligo's sequence**
cDNA Rab38/32 EcoRI Forward	5'CCGGAATTCCGGTTATCTGCAACAACCGCTCTTTTC 3'
cDNA Rab38/32 NotI Reverse	5'AAGGAAAAAAGCGGCCGCTTTAATGACCGAACAAAGAAAGGA 3'

2.31 Preparation of constructs used for Ci-Rab38/32/rab-rp1/ltd mutated protein

The pTyr>Rab38^{G19V} and pTyr>Rab38/32$^{G19V+T23N}$ constructs have been prepared using the QuikChange® Site-Directed Mutagenesis Kit "Stratagene" from the pTyr1/2a>Rab38/32 construct. The dominant negative form (Rab38G19V) has been obtained introducing a single nucleotide substitution in codon 19 (GGT to GTT) that results in a Gly to Val substitution in *Ciona* switch I region of GTPase binding pocket, while to obtain the double mutant I introduced a double nucleotide substitution (ACG to AAT) in 4 nucleotide from the previous mutation, at position 23 of *Ciona* Rab38/32 protein sequence. The mutagenic oligonucleotides used are listed in Table 5, ordered as synthetic single strand oligonucleotides by the Primm Company. According to manufacturer's instruction, these mutated oligonucleotides have been used for a PCR reaction using as template the pTyr1/2a>Rab38/32 construct. The PCR reaction has been carried out using the PFU DNA polymerase and these cycling parameters: after a single denaturation step at 95°C for 30 seconds, 18 amplification cycles were performed as follows: denaturation at 95°C for 30 seconds, annealing at 55°C for 1 minute, extension at 68°C for 7 minutes. The presence of amplified product has been checked loading 10 µl of the reaction on 1% agarose gel. At this point to eliminate the FG wild type plasmid, used as template, the mixture has been digested at 37°C for 1.5 hr with 1 µl of the DpnI restriction enzyme (10 U/µl). This enzyme digests only the methylated supercoiled dsDNA used as template but not the newly synthesized mutated and unmethylated DNA. After the digestion, 2 µl of the reaction has been transformed and grown as previously described (Par. 2.11). Ten clones have been selected and after growth, the isolated plasmids DNAs have been sequenced to check the presence of mutations.

pTyrp1/2a>mCherry

To obtain the pTyrp1/2a>mCherry construct, I double digested the plasmid pTyrp1/2a>dnFGFR with NotI and EcoRI to eliminate the coding sequence for dnFGFRDN and purified the band with pTyrp1/2a fragment by gel extraction. This

sequence has been inserted in the vector pTyr>mCherry (already available in the lab), that has been previously digested with NotI and EcoRI to eliminate the pTyr, and purified by gel extraction. The ligation reaction has been conducted using an optimal proportion of 1 : 3, vector : insert.

Table 5: Oligonucleotides used for mutagenesis	
Mutated nucleotides are indicated in bold character and underlighted in red	
Oligo's name	**Oligo's sequence**
Rab38 G19V Forward	5'gttcttgtgatcggcgagttag**t**tgttgggaagacgagcgtcatt3'
Rab38 G19V Reverse	5'aatgacgctcgtcttcccaac**a**actaactcgccgatcacaagaac3'
Rab38 T23N Forward	5'cgagttagttgttgggaaga**at**agcgtcattaaaagatat3'
Rab38 T23N Reverse	5'atatcttttaatgacgct**at**tcttcccaacaactaactcg3'

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Microarray experimental design

Due to the complexity of gene expression in a biological sample, the experimental design has a critical importance for the statistical and biological validation of microarray data. The problem of designing a microarray experiment can be composed into three main elements: first, replication of biological samples is essential in order to draw conclusions that are valid beyond the scope of the particular samples that were assayed, second, technical replicates increase precision and provide a basis for testing differences within treatment groups, third, duplication of spotted clones on the microarray slides increases precision and provides quality control and robustness to the experiment .

For my microarray experimental design, I induced a molecular perturbation of the endogenous FGF cascade using two molecular constructs that resulted fundamental to demonstrate that FGF-MAPK signaling is necessary for pigment cell induction in *C. intestinalis*: a dominant negative form of the sole Ci-FGF Receptor consisting of a mutated receptor lacking the intracellular PTK domain (dnFGFR) and a constitutively active form of the Ci-Ets1/2, a transcriptional effector of FGF/MAP Kinase cascade, consisting of the Ets1/2 DNA-binding domain fused to the VP16 trans-activation domain (Davidson et al., 2006). I will subsequently refer to the dnFGFR as inhibited condition and the condition as EtsVP16 as hyper activated respectively.

To drive a targeted manipulations of the FGF signalling cascade to the a9.49 pigment cell precursors (PCPs), I used the *cis*-regulatory region from *Tyrp1/2a* (Esposito et al., 2012), which starts to be active at late gastrula stage, almost contemporarily with the requirement of the MAPK cascade to direct pigment cell fate induction. I considered to isolate PCP cells at two differential developmental stages: at 8 hour post fertilization (hpf) at ~16°C, approximately neurula stage, and 12 hpf at ~16°C, tailbud stage.

The 8hpf stage corresponds to FGF-mediated early induction because, at the this stage, division of the two a9.49 cell pairs already occurred and gave rise to two a10.97 cell pairs, the pigment cell precursors, and the a10.98 cell pair. At middle tailbud stage the four PCPs intercalate at the dorsal midline, aligning in a single row along the anterior-posterior axis of the animal. At this 12hpf stage, pigment cells are fate restricted.

To determine how FGF signaling induces pigment cell specification, I employed a method for lineage specific transcription profiling using Fluorescence Activated Cell Sorting (FACS) and microarray analysis. PCPs-lineage cells were sorted based on their GFP

fluorescence. In order to visualize fluorescent PCPs at neurula (8hpf) and tailbud stage (12hpf), I performed transgenesis experiments *via* electroporation of the construct *pTyrp1/2a*>GFP. I could visualize GFP expressing cells just at 12hpf embryos that show the expression of the reporter gene in all the four PCP cells aligned along the dorsal midline of the developing neural tube. I didn't observe GFP[+] cells at neurula stage, because *Tyrp1/2a* starts to be expressed at late gastrula stage and the *cis*-regulatory region used could not induce a sufficient accumulation of the reporter fluorescent protein in PCPs at 8hpf embryos. In order to visualize GFP[+] cells at neurula stage, I create another construct cloning the *cis*-regulatory DNA of *Tyrp1/2a* upstream two coding sequences in frame for GFP, that I called *pTyrp1/2a*>2xGFP; using this construct, I could visualize the fluorescent reporter expression in the four PCP cells also at 8hpf (Fig. 3.1).
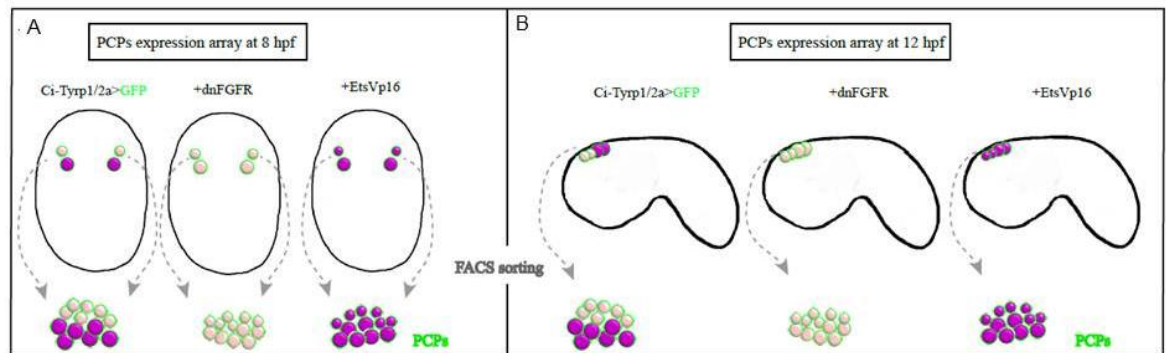
Fig.3.1 Schematic representation illustrating the two developmental stages and the three conditions used for microarray experiments (A,B). PCPs-lineage cells have been isolated from 8 hours post-fertilization (hpf)(A) and at 12 hpf (B); they were sorted based on their GFP fluorescence in embryos electroporated with Ci-Tyrp1/2a>GFP (control sample), Ci-Tyrp1/2a >dnFGFR (Inhibited condition) and Ci-Tyrp1/2a>EtsVp16. (Hyper activated condition).

Ciona promoter regions are often activated in another territory, the mesenchyme; using FACS, this cellular population can be collect together aspecifically with the cells of my interest; to avoid a cellular contamination of the mesodermal cells, *MyoD cis*-regulatory sequence was used to drive the expression of another reporter gene, Yellow Fluorescent Protein (YFP), in the mesenchyme cells in order to counter select this cellular population from the GFP[+] sorted cell (Fig 3.2).
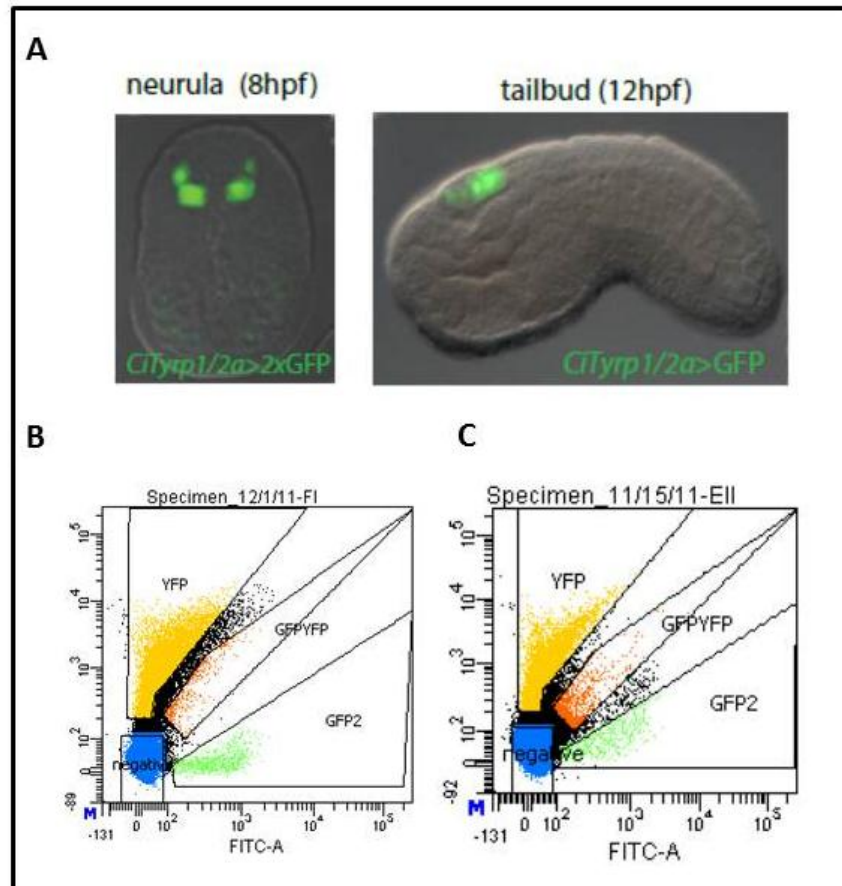
Fig.3.2 FACS plots(A-C): Pigmented precursors GFP expressing cells: A) Embryos expressing pTyrp1/2a>2xGFP at neurula stage (8hpf) and pTyrp1/2a>GFP at tailbud stage (12hpf). B-C) FACS plot referred to pTyrp1/2a>2xGFP (B), pTyrp1/2a>GFP (C) and pMyoD>YFP-expressing cells are distinguished by their green or yellow fluorescence, respectively. The GFP-only quadrant (green box) was used to sort Pigmented Cell Precursors using FACS.

Furthermore to electroporate the same amount of DNA in each condition, for the control samples I used a construct that does not influence the signalling and has a similar dimension of the other used for the perturbation, such as *ptyrp1/2a*>EtsVp16 and *ptyrp1/2a*>dnFGFR; this construct is composed by the coding sequence of β-Galactosidase cloned downstream the *pTyrp1/2a cis*-regulatory DNA (*pTypr1*/2a>LacZ).

Transgenesis experiments have been performed using a solution containing the DNA construct of interest:

- *pTyrp1*/2a>2xGFP as control to isolate the pigment cell precursors at the neurula stage
- *pTyrp1*/2a>GFP as control to isolate the pigment cell precursors at the middle tailbud stage
- *pMyoD*>YFP to counter select cells derived from mesenchyme
-*pTyrp1*/2a>dnFGFR for the inhibited condition
-*pTypr1*/2a>EtsVP16 for the hyper activation condition
- *pTypr1*/2a>LacZ for the control samples.

For each targeted perturbations and control samples, I collected four biological replicates to ensure rigorous statistical analysis of the results obtained, performing transgenesis experiment *via* triple electroporation of DNAs (Fig 3.3).
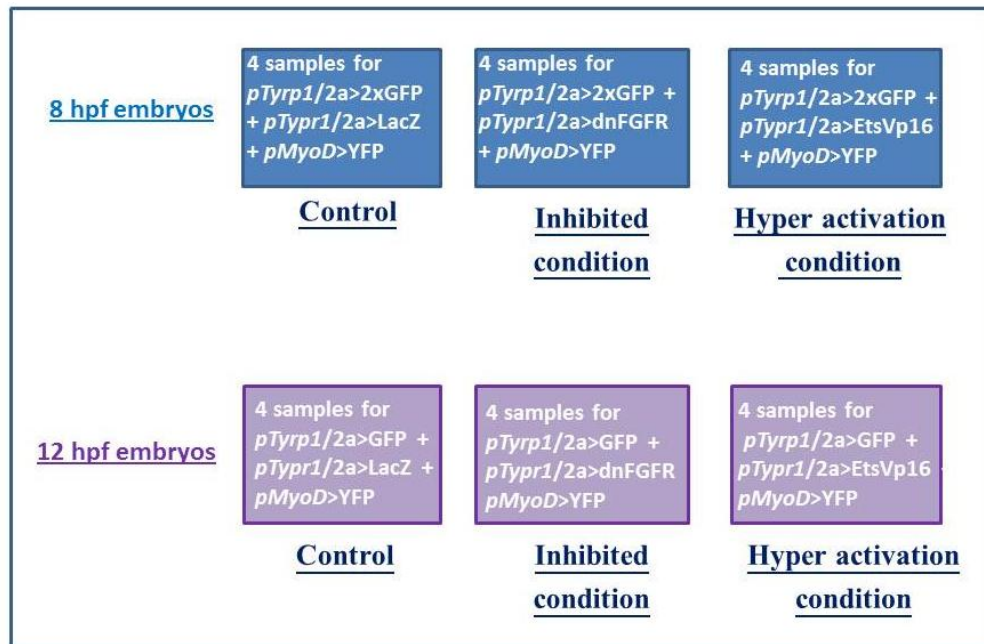
Fig.3.3 Scheme for triple electroporation with the DNA combinations used to prepare embryos for microarray experiments: for each targeted perturbations (pTyrp1/2a>dnFGFR and pTyrp1/2a>EtsVp16) and control samples(pTyrp1/2a>LacZ), four biological replicates have been collected at 8 and 12 hpf embryos. All the samples have been prepared by a triple electroporation of DNAs, with pMyoD>YFP to specifically mark mesenchyme cells and with pTyrp1/2a>2xGFP to induce GFP expression at 8hpf samples while pTyrp1/2a>GFP for 12hpf samples.

Sorted cells from embryos electroporated were collected for RNA extraction and RNA quality and integrity were analysed by BioAnalyzer (Agilent Technologies) that provides a RNA Integrity Number (RIN), calculated by an algorithm that assigns a 1 to 10 RIN score, where level 10 RNA is completely intact values. To ensure a good quality samples for microarray experiments, I used RNA samples with RIN score comprising 9.5 and 10 score (Fig 3.4).
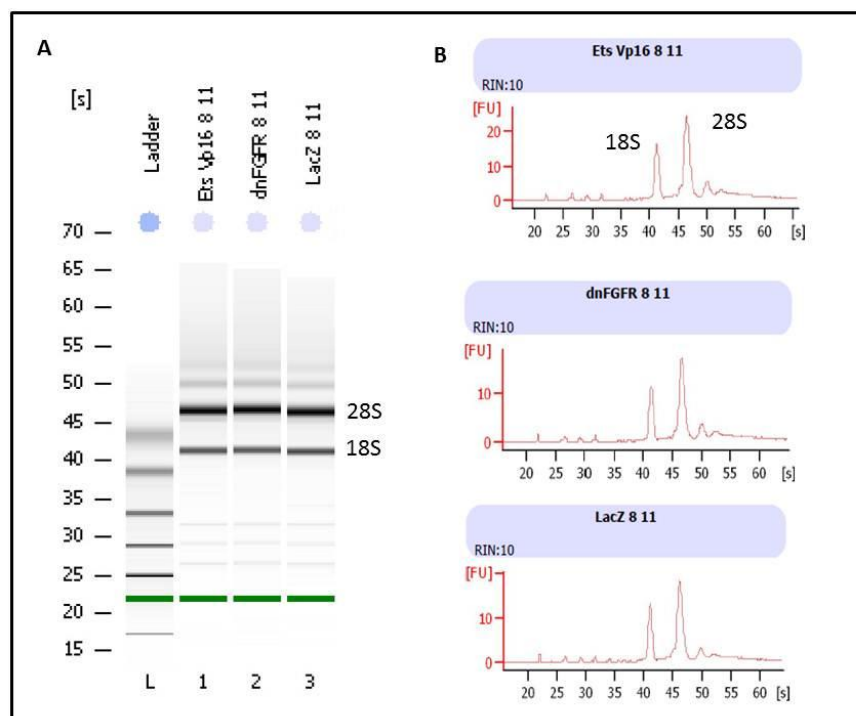
Fig. 3.4 Determination of RNA integrity and quality (A,B): A) Bioanalyzer gel-like image of total RNA samples run on an Agilent 2100 Bioanalyzer RNA 6000 Pico LabChip. L, Ladder/marker; 1-EtsVp16, 2- dnFGFR, 3-LacZ samples; B) Electropherogram of Fluorescent trace with RNA 18S and 28S peaks. The RNA integrity Numbers (RIN) scores are all equal to 10.

3.2 Whole Genome Ciona intestinalis microarray annotation

The whole genome microarray design was based on previous annotations of the *Ciona intestinalis* genome assembly, version 1.0 (Dehal et al., 2002). A total of 19,682 Kyotograil2005 gene models (Satou et al., 2005) and 6,913 full-length cDNA were used together with 711,961 ESTs (including ~650,000 initial ESTs and additional clones that were made available to us prior to publication by Yutaka Satou (Satou et al., 2006)). These transcript models were aligned to the genome assembly by BLAT. The regions where these transcript sequences aligned were treated as candidate transcribed regions of the *Ciona* intestinalis genome. A total of 30,880 candidate transcribed regions were obtained. This set was then compared with ENSEMBL gene models and GenBank data, which led to the addition of 357 transcript models. The final set was made of 31,237 candidate transcribed regions. The complete data set was submitted to Affymetrix for probe design. Nucleotide sequences used to design the probes for each probesets by Affymetrix are stored in the Affymetrix library file (CINT06a520380F.sif). The total GeneChip microarray contains 30,970 distinct probe sets (each made of 16 individual probes, except for the tiling probe sets). After normalization process, all 16 probes corresponding to one transcript are summarized giving an average value referring to a single transcript that is further used for the statistical analysis. 29,833 of these probe sets (96.3%) correspond to 21,621 independent transcripts models (modified KyotoGrail2005 models and EST clones) (Christiaen et al., 2008). Because *Ciona intestinalis* GeneChips are custom microarray, we created a probe annotation file to associate every probeset to a specific gene and its functions. In order to create the annotation file, the nucleotide sequences of probes are retrieved from Affymetrix library file (CINT06a520380F.sif) while the transcript models and annotations were downloaded from the ANISEED database (Tassy et al., 2010) on March 2011: 1) JGI version 1, 2) KYOTOGRAIL2005, 3) KH and 4) ENSEMBL. Transcript models and annotations were parsed and joined together to form a unique collection. Probeset sequences were then aligned against the collection of transcript models by using BLAST+ (Camacho et al., 2009) with the megablast algorithm option. Results were filtered in order to select exclusively hits with coverage >75% and identity >90%. Based on the selected hits, we associated each probeset to the corresponding transcript models and annotations (Fig. 3.5). To perform such analysis we developed custom perl scripts making extensive use of the Bioperl toolkit (Stajich et al., 2002). Finally, a MySQL database containing all the information downloaded and generated was used to collect and manage the data.
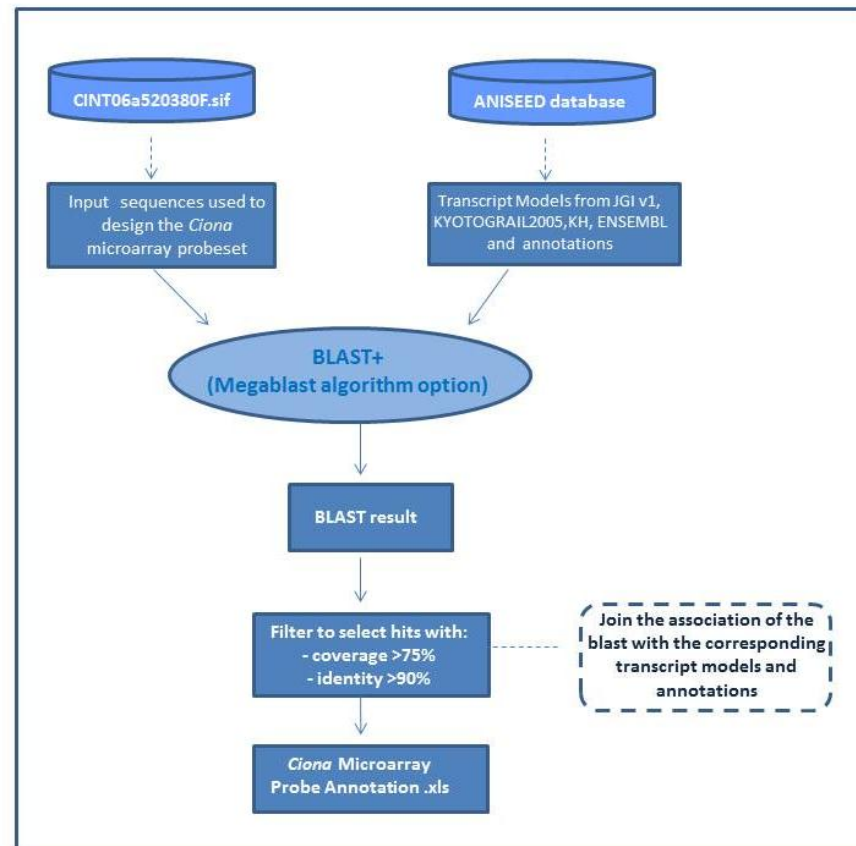
Fig. 3.5 Pipeline for whole genome *Ciona intestinalis* microarray annotation.

3.3 Microarray Data Analysis

Microarray technologies generate huge amount of data. A typical microarray data set includes expression levels for thousands of genes across several conditions.

Converting the data provided in a microarray dataset into meaningful biological information involves several steps including: i) normalization ii) selection of differentially expressed genes using comparison statistics iii) identification and partition of expression patterns based on up and down regulation, magnitude and clustering of genes iv) summarizing the identified gene expression patterns using gene annotation tools. Specifically, microarray data analysis involves quality control, normalization, statistical analysis, clustering and classification.

For this thesis project, microarray data analysis has been done using GeneSpring GX 12, a popular gene expression analysis software that provides powerful, accessible statistical tools for fast visualization and analysis of expression and genomic structural variation data. The first operation in GeneSpring software to analyze microarray data is to create a technology that is defined as the species specific genome printed on microarray chip including biological and other information about the probes, *i.e.* Entrez gene ID, GO accession, Gene Symbol etc. The data processing using GeneSpring software depends on the technology; for Affymetrix Expression Data, this processing includes:

Raw signal values: in Affymetrix Expression experiment, the term "raw" signal values refer to the linear data after summarization algorithms; visualization of raw data is an essential part of assessing data quality, choosing a normalization method and estimating the effectiveness of the normalization.

Normalized signal values: using GeneSpring GX 12, the software performs as default mode the normalization processing using RMA method (Background correction, Normalization, Summarization); subsequently a $\log_2$ transformation and a baseline transformation is performed, that correspond to a linear values conversion in $\log_2$ values and to a "Baseline to median" transformation, where for each probe the median of the $\log_2$ summarized values derived from all the samples is calculated and subtracted from each of the samples (Fig. 3.6) .
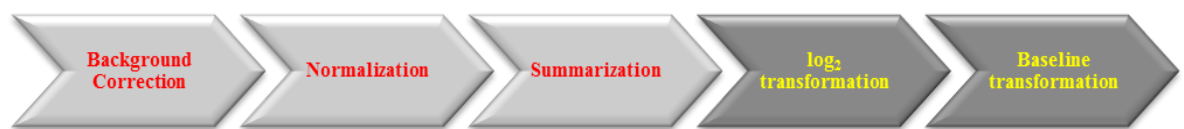
Fig. 3.6 Order of processes for normalization in GeneSpring GX 12

3.3.1 Quality Assessment of Affymetrix Genechip Data

There are a number of useful tools implemented to assess the quality of GeneChip data. In order to ascertain data quality after all data have been collected and check which samples are ambiguous and remove from the analysis, I performed Quality Control analysis (QC) that includes Box whisker plot of normalized microarray data, scatter plots for biological replicate comparison, Principal Component Analysis (PCA), Correlation analysis, Affymetrix hybridization controls and filtering probeset by expression values (see Appendix I).

3.3.2 Statistical analysis

The fundamental goal of microarray experiments is to identify biological processes or pathways that consistently display differential expression between groups of samples. Differential expression analysis of microarray data is fraught with many classical statistical issues, such as appropriate test statistics, replicate structure, sample size, outlier detection and statistical significance of results. The original and simplest approach to identifying differentially expressed genes was to use a fold change criteria; selecting cutoff was something of an ad-hoc procedure; a 1.5-fold change was however thought as being a suitable cutoff. There are a number of statistical tests available that can be applied to assess differential expression between populations of microarray data, such as the t-test, which

can be used to assess the statistical probability that, given the number of samples available, the true expression levels for a given gene differ in the overall populations.

After assessed the quality of microarray dataset, I processed to identify statistically significant genes using a t-test and one-way ANOVA on the normalized data selecting significant differential transcript with a p-value < 0.05. Furthermore, a fold change analysis of significance was performed to address the magnitude of change of statistically significant genes.

Regarding the t-test analysis, I chose the unpaired t-test because the four biological replicates for each condition have been obtained with RNA extraction performed on embryos collected in different days and from different batch of animals. Each statistical tests reports the probability of seeing the observed test score by chance under the null hypothesis that there is no difference in expression related to the condition tested. I considered the comparison among all the three conditions (dnFGFR, EtsVp16 and Control) respectively at the two time point condition, 8 and 12 hpf (Fig. 3.7 and Table1).
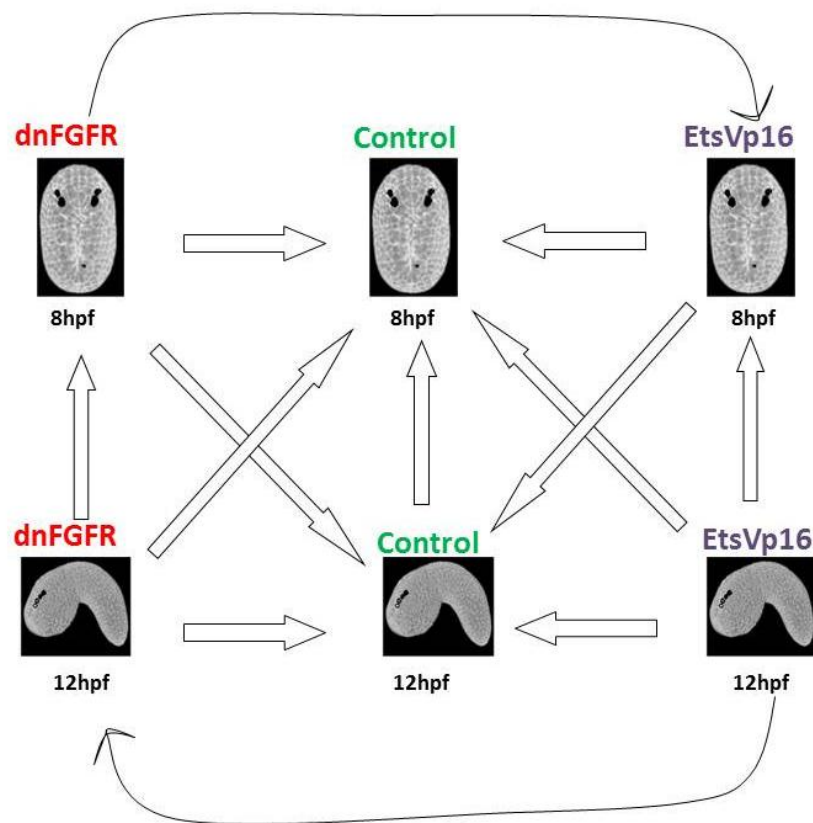
Fig. 3.7 Representative scheme for all the sample combinations applied for the statistical analysis. Embryos have been represented at the two developmental stages (8 and 12 hpf) treated for the three condition (control, dnFGFR, EtsVp16); PCP cells are represented by black spots.

Table 1: Statistical analysis on microarray data set			
Combination of samples analyzed	**Statistical Test**	**# transcripts filtered by p-value < 0.05**	**# transcripts with Fold Change > 1.5**
dnFGFR vs EtsVp16 12hpf	t-test	2886	1626
dnFGFR vs EtsVp16 8hpf	t-test	720	169
Ctrl 8hpf vs Ctrl 12hpf	t-test	13310	9191
dnFGFR vs Ctrl 12hpf	t-test	2246	958
dnFGFR vs Ctrl 8hpf	t-test	539	144
EtsVp16 vs Ctrl 12hpf	t-test	1809	905
EtsVp16 vs Ctrl 8hpf	t-test	260	1
dnFGFR 12hpf vs dnFGFR 8hpf	t-test	13154	8350
EtsVp16 8hpf vs EtsVp16 12hpf	t-test	13292	9116
dnFGFR vs EtsVp16 vs Ctrl 12hpf	ANOVA	2875	1526
dnFGFR vs EtsVp16 vs Ctrl 8hpf	ANOVA	519	137

Notably analysing the distribution of fold change ratios and their relationships with p-values comparing the two conditions, dnFGFR and EtsVp16, to Control samples at the two time point stages, I found a robust discrepancy in differential expressed genes displaying significant changes at 8 hpf (114 transcripts for dnFGFR and 1 for EtsVp16) comparison to 12 hpf samples (958 transcripts for dnFGFR and 905 for Etsvp16); this data confirms that FGF signaling cascade have a stronger effect on the induction of pigment cell precursor cells at tailbud stage than at neurula stage; indeed during the primary FGF signaling induction at neurula stage (8hpf) genes are more likely to be maintained in a stable off state while most of genes are subsequently activated by FGF cascade at tailbud stage. Moreover statistical analysis of the microarray dataset evidences that there is a less gene expression changes upon EtsVp16 at 8hpf considering that at ~7.5hpf, the *cis*-regulatory region of *Tyrp1/2a* starts to be active leading the expression of both conditions used to interfere; at 8 hpf, the early expression of dnFGFR has a stronger impact on genes regulated by FGF signalling cascade because the mutation of a receptor leads to an immediate block of the cascade whereas a constitutive active form of a transcription factor needs more time to be accumulated and starts working; considering these considerations, I

found consistent that there is a discrepancy in differential expressed genes when the two conditions are expressed at 8 hpf.

In the following paragraph I describe in details the statistical analysis that allow me to select new candidate genes showing a significant differential expression and that I subsequently validated.

3.3.3 Identification of differential expressed genes comparing dnFGFR to Control samples both at 12 hpf and 8 hpf

To identify new candidate downstream FGF signaling, I firstly compared gene expression profiles of the dnFGFR and Control samples at 12hpf. The criteria for significant difference were set as p-value <0.05 and fold change > 1.5.

Of the total 23822 transcripts that have been normalized and filtered upon their signal intensity values, there were 958 transcripts that were significantly different between dnFGFR and control, including 531 up-regulated and 427 down-regulated.

A Volcano plot has been used to report the differential expression analysis result: the *x*-axis is $\log_2$ ratio of transcript expression levels between dnFGFR and Control, while *y*-axis is p-value based on $-\log_{10}$. The colored dots represent the differentially expressed transcripts based on $p < 0.05$, while the grey dots represent transcripts that have not passed p-value cut off. The light blue dots represent transcripts that have a $\log_2$ ratio higher 0.6 that I further consider to biologically validate, while the red dots have a $\log_2$ ratio lower 0.6 (Fig. 3.8).
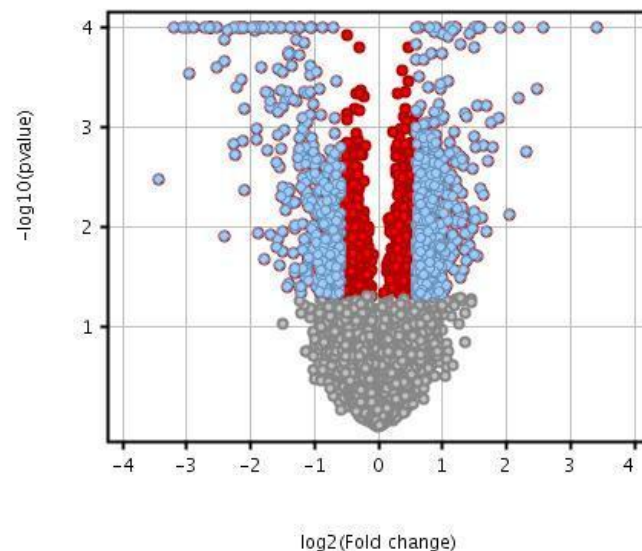
Fig. 3.8 **Significant differentially expressed transcripts comparing dnFGFR to Control samples at 12 hpf.** Volcano plot of the expressed transcripts, with differential expression magnitude plotted on the *x*-axis (dnFGFR/control on the log2 scale (log2 fold change)) and the significance of differential expression (negative log10-transformed p values) plotted on the *y*-axis (n = 958 subjects). Grey dots represent not differentially expressed; red and blue dots represent differential expressed transcripts, with p-value<0.05 ; only blue dots have FC >1.5 .

In order to reduce the probability that any particular significant finding could be a false positive, I applied a more stringent criteria adjusting p-values [adj.-p] < 0.05 corrected for multiple testing by the Benjamini and Hochberg algorithm (Benjamini Y. et al., 2001); as excepted, I observed a robust decreased in significant transcripts selected with a total of 48 that were significantly regulated when comparing dnFGFR and control samples; in particular as shown in the Volcano plot, 15 transcripts were up-regulated, 33 down-regulated (Fig. 3.9).
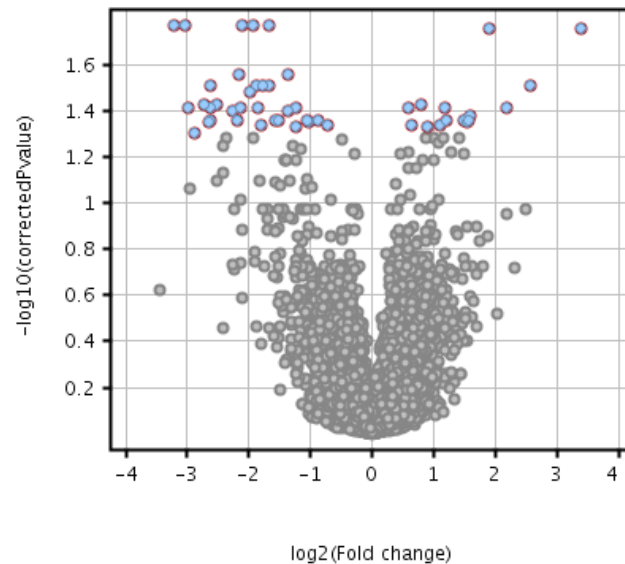
Fig. 3.9 Corrected significant differentially expressed transcripts comparing dnFGFR to Control samples at 12 hpf. Volcano plot of the expressed transcripts, with differential expression magnitude plotted on the *x*-axis (dnFGFR/control on the log2 scale (log2 fold change)) and the significance of differential expression (negative log10-transformed Corrected p values) plotted on the y-axis (n = 958 subjects). Grey dots represent not differentially expressed; blue dots represent differential expressed transcripts, with p-value<0.05 and FC >1.5 .

Although most of genes selected without Multiple Testing Correction were excluded, know marker genes for PCPs, as *Ci-BMP5/7*, *Ci-Tcf*, are still present in the list of differentially expressed transcripts.

I subsequently compared gene expression profiles of the dnFGFR and control samples at 8 hpf; the criteria for significant difference were set as for the previously analysis with a p-value <0.05 and fold change > 1.5. Of the total 23957 transcripts that have been normalized and filtered upon their signal intensity values, there were 144 transcripts that were significantly different between dnFGFR and control, including 124 up-regulated and 20 down-regulated as shown in Volcano plot (Fig. 3.10).
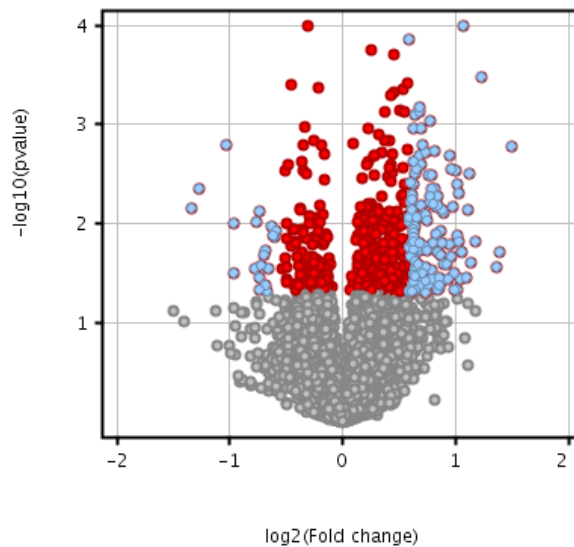
Fig. 3.10 Significant differentially expressed transcripts comparing dnFGFR to Control samples at 8 hpf. Volcano plot of the expressed transcripts, with differential expression magnitude plotted on the *x*-axis (dnFGFR/control on the log2 scale (log2 fold change)) and the significance of differential expression (negative log10-transformed p values) plotted on the y-axis (n = 114 subjects). Grey dots represent not differentially expressed; red and blue dots represent differential expressed transcripts, with p-value<0.05 ; only blue dots have FC >1.5 .

Also in this case I applied to adjust p-values [adj.-p] < 0.05 corrected for multiple testing by the Benjamini and Hochberg algorithm, but no significant transcript passed this more stringent criteria.

I consider the 8 hpf time point condition as a FGF-mediated early induction on PCP cells; based on the detection of driven reports, the *pTyrp1/2a cis*-regulatory region starts to be active at late gastrula stage (7.5 hpf); so the mutated form of FGF receptor, dnFGFR, start to be produced at significant levels by the beginning of 8 hpf. Considering this time expression, I excepted that a lower number of genes are influenced by targeted FGF perturbation at 8 hpf because dnFGFR expression will have less impact on gene regulation until FGF mediated expression first occurs, at ~ 7.5hpf. Thus it's consistent that I found this discrepancy in number of gene differential expressions when comparing dnFGFR and control samples at 8 and 12 hpf.

3.3.4 Identification of differential expressed genes comparing EtsVp16 to Control samples both at 12 hpf and 8 hpf

I next proceed analysing gene expression profiles of the EtsVp16 and control samples at 12 hpf; the criteria for significant difference were set as for the previously analysis with a p-value <0.05 and fold change > 1.5.

Of the total 23813 transcripts that have been normalized and filtered upon their signal intensity values, there were 905 transcripts that were significantly different between

EtsVp16 and control, including 698 up-regulated and 207 down-regulated, as shown in Volcano plot (Fig. 3.11).
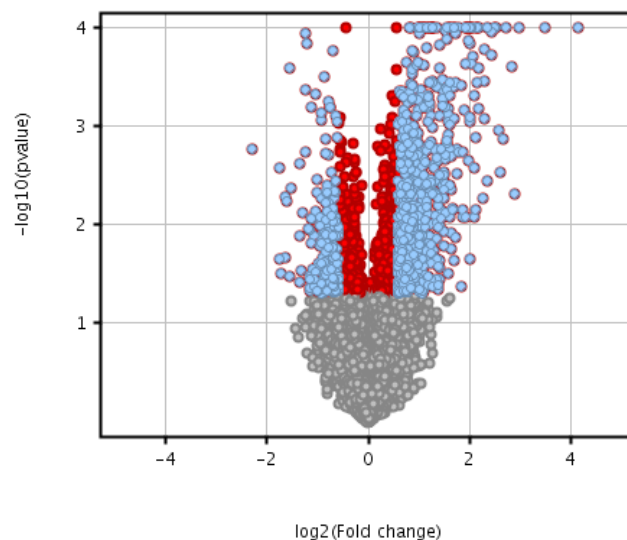
Fig. 3.11 Significant differentially expressed transcripts comparing EtsVp16 to Control samples at 12 hpf. Volcano plot of the expressed transcripts, with differential expression magnitude plotted on the *x*-axis (EtsVp16/control on the log2 scale (log2 fold change)) and the significance of differential expression (negative log10-transformed p values) plotted on the y-axis (n = 905 subjects). Grey dots represent not differentially expressed; red and blue dots represent differential expressed transcripts, with p-value<0.05 ; only blue dots have a FC >1.5 .

Also in this case I applied to adjust p-values [adj.-p] < 0.05 corrected for multiple testing by the Benjamini and Hochberg algorithm to reduce false positive in differential significant finding; using this adjusting for multiple test correction, I selected 29 significantly regulated, all up-regulated, as shown in the Volcano plot below. In this case, no known marker gene was present in the gene list (Fig. 3.12).
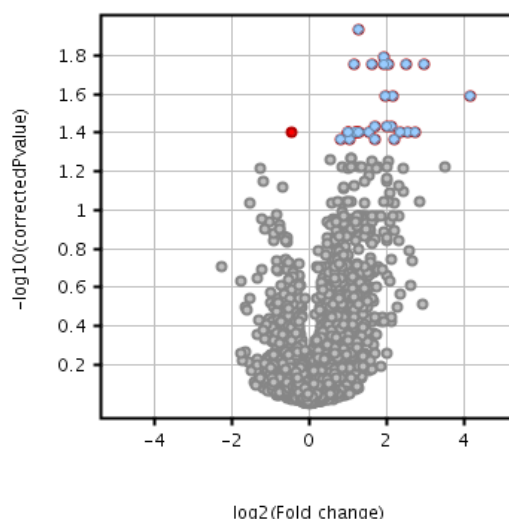
Fig. 3.12 Corrected significant differentially expressed transcripts comparing EtsVp16 to Control samples at 12 hpf. Volcano plot of the expressed transcripts, with differential expression magnitude plotted on the *x*-axis (EtsVp16/control on the log2 scale (log2 fold change)) and the significance of differential expression (negative log10-transformed Corrected p values) plotted on the y-axis (n = 905 subjects). Grey dots represent not differentially expressed; red and blue dots represent differential expressed transcripts, with p-value<0.05 ; only blue dots have a FC >1.5 .

I subsequently performed the same analysis comparing EtsVp16 and control samples at 8 hpf; of the total 23963 transcripts that have been normalized and filtered upon their signal intensity values, there were 260 transcripts that were significantly different between dnFGFR and control, but neither of them show a fold change ratio higher 1.5; also in this case, applying the adjusting p-value correction no transcripts passed the cut-off criteria.

3.3.5 Identification of differential expressed genes comparing Control samples at 12 hpf to Control samples 8 hpf .

In order to identify genes that are regulated on time in PCP cells during FGF induction, I compared gene expression profiles of the control samples at 12hpf and control samples at 8 hpf. I applied a t-test analysis using the same criteria of previously tests (p-value < 0.05 and fold change > 1.5).

In this case, a higher number of transcripts passed the cut-off criteria; of the total 24312 transcripts that have been normalized and filtered upon their signal intensity values, there were 11782 transcripts that were significantly different between the two control samples (Fig. 3.13).
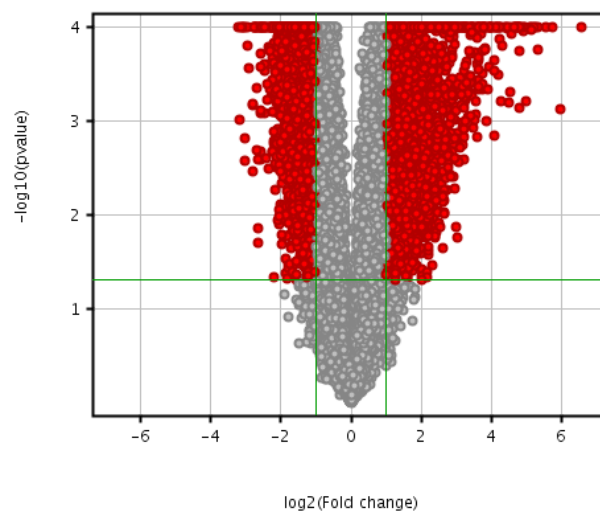
Fig. 3.13 Significant differentially expressed transcripts comparing Control at 12hpf to Control samples at 8 hpf. Volcano plot of the expressed transcripts, with differential expression magnitude plotted on the *x*-axis (control_12/control_8 on the log2 scale (log2 fold change)) and the significance of differential expression (negative log10-transformed p values) plotted on the y-axis (n = 11782 subjects). Grey dots represent not differentially expressed; red dots represent differential expressed transcripts, with p-value<0.05 and with FC >1.5 ; the horizontal green line represent p-value cut-off, the two vertical green lines are for FC cut off.

In order to reduce the false positive transcripts, I applied for Bonferroni correction to adjust p-values [adj.-p] < 0.05, a more stringent Multiple Testing Correction Method than the Benjamini and Hochberg algorithm. As excepted using a more stringent algorithm, I observed a robust decreased in significant transcripts selected with a total of 45 significantly regulated transcripts, resulting in 22 up-regulated and 23 down-regulated, as

shown in the Volcano plot below. The list of genes that show a significant differential expression between the two time point conditions was used to identify candidates that are regulated in time during FGF signalling induction.

3.4 Gene Ontology analysis on differential expressed genes

The *Ciona* GO annotations for each probeset were extracted from the Aniseed database as already described. The analysis have been performed on transcripts that have a significant differential expression when comparing dnFGFR to control samples at 12hpf. The groups of differentially probesets were compared with all the probesets annotated on the array considering the proportion of probesets associated to each GO class in the 2 groups. Significance was assessed using the Fisher exact text and the p-values were corrected using the Benjamini and Hochberg method. The GO analysis result is reported in Appendix III and Fig 3.14.
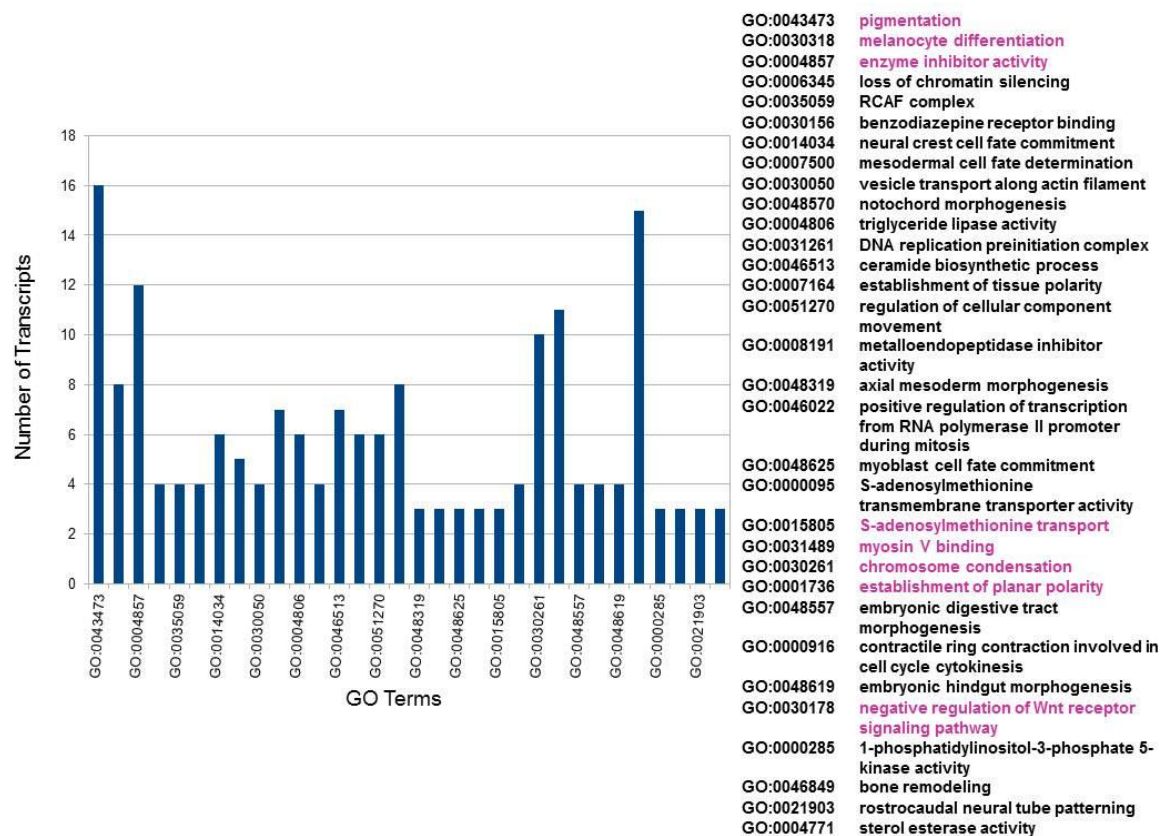
Fig. 3.14 GO analysis on differentially expressed transcripts comparing dnFGFR to Control at 12 hpf.
Bar chart for GO analysis, on *y*-axis, number of transcripts is represented; on *x*-axis the most significant represented GO terms, ranked for significancy representation, upon corrected p-value values. The legend on the right shows the GO term and class; the most representative are evidenced in pink.

It's interesting to note that the two more significant GO classes identified are referred to "pigmentation" (GO:0043473) and "melanocyte differentiation" (GO:0030318); in the pigmentation class, 16 transcripts have been selected, while in the melanocyte

differentiation class, 8 transcripts; all of these have not be characterized in *Ciona* except for *Ci-Mitf*, and no evidences are reported about their involvement in *Ciona* pigmented cell formation. Among these 16, there is the ortholog of mouse *HPS1* (*Hermansky-Pudlak syndrome 1*), a gene encoding a protein that play an important role in LROs biogenesis, organelles comprising melanosomes, platelet dense granules, and lysosomes; mutations in this gene are associated with Hermansky-Pudlak syndrome type 1; thus they all can be considered as good candidates to find new FGF/MAPK/Ets target genes involved in *Ciona* pigment cell development.

Furthermore, in the most significant GO classes, there is "negative regulation of Wnt receptor signaling pathway" (GO:0030178), with 15 transcripts selected; among these , there is Six3/6, that is a repressor of Wnt receptor and is necessary to inhibit the expression of Wnt signaling in the anterior neuroectoderm for a correct vertebrate forebrain development (Zhu et al., 2002); in the list are present also Lrp4, a Wnt co-receptor molecule (Ohazama et al., 2008) and Dickkopf , another potent soluble Wnt inhibitor that binds to Lrp co-receptor (Choi et al., 2009).

Wnt signaling has a fundamental role in the *Ciona* pigment cell differentiation (Squarzoni et al., 2011); thus all of the transcripts selected in this GO class represent good candidates to be biologically validated and studied to understand their role in FGF and Wnt signaling communication.

3.5 Reverse engineering of tissue-specific gene co-regulation networks and neighbors analysis

This work has been done in collaboration with Gennaro Gambardella from di Bernardo's group (Telethon Institute of Genetics and Medicine-TIGEM, Naples). In order to create the reverse engineering of tissue-specific gene co-regulation network, we used all the 24 samples normalized as previously described using RMA (Robust Multichip Average) algorithm (Par. 3.3). For the reconstruction of the co-regulatory network, we computed the Spearman Correlation Coefficient (see Par. 2.22) for each pair of probes present in the chip, obtaining a final correlation matrix of dimension (30'974 x 30'974),

In order to control the number of False Positives due to the multiple hypotheses test problem, we estimated the degrees of freedom of the t-test distribution from the data by fitting the parameters of a Student's t-location-scale distribution to the t statistics computed for all the probe pairs. We estimated the parameters by minimizing the squared error between the theoretical and the empirical distribution and only probes with a p-value less than 0.01 have been considered for the following analysis (for more details on the creation

of co-regulated networks see the Chapter 3 of Gambardella's thesis on "Identification of transcriptional and post-translational regulatory networks from gene expression profile: an information-theoretic approach").

After the transcriptional co-regulatory network has been created, we subsequently continued with neighbors analysis. The definition of a neighbor is the following: let a probe *p*, the neighbors of first level of *p* are all the other probes in the regulatory network showing a significant value of Spearman Correlation Coefficient (SCC) according to our setting, p-value < 0.01.

We analyzed the neighbors of level one for the probes, showed in Table 2; the selection of these transcripts is based on their involvement in pigment cell formation and induction, in *Ciona* (Squarzoni et al.,2011) and vertebrate melanocytes (Ohbayashi and Fukuda, 2012).

Table 2 Probes selected for neighbors analysis	
Probe set IDs	**Ciona Name**
MKG.166.29.1_at	Ci-Tyrosinase
MKG.61.27.s1_s_at	Ci-Tcf/LEF
MKG.13.94.1_at	Ci-Tyrp1/2a
MKG.42.3.1_at	Ci-Tyrp1/2b
MKG.1330.3.1_s_at	Ci-Rab38/32

In Appendix IV, the neighbors analysis results for these ribo probes are reported in Table 1-5, with all the probe set IDs ranked for the SCC values. Interestingly, I observed that the most representative ribo probes with higher SCC values, are associated to genes involved in melanization, such as *Ci-Tyrosinase*, *Ci-Tyrp1/2a*, *Ci-Tyrp1/2b*, *Ci-Tcf/LEF1*, that could be considered as a positive control for the neighbors analysis. Moreover I observed that several genes, such as *Ci-scl45a2*, *Ci-Exosc7*, *Ci-Rab3GAP* involved in melanosomal logistic, including biosynthesis, maturation and movement of melanosomes in the cells, are present in the lists. Starting from microarray gene expression data, this reverse engineered approach turned out to be a potential method to identify new candidates that are co-expressed with melanogenic genes and in turn implicated in pigment cell formation; because they have been recovered computationally, further biological validations are necessary to demonstrate their involvement in the gene regulatory network controlling pigmentation in *Ciona* sensory organs.

3.6 *Training set data and global analysis for microarray profiling*

The microarray data were first evaluated using expression and significancy values for known PCPs genes, that include the melanogenic genes, *Ci-Tyrosinase, Ci-Tyrp1/2a, Ci-Tyrp1/2b* and *Ci-BMP5/7, Ci-Tcf/LEF, Ci-MITF*. The main expectation is that the inhibited

condition at 8 and 12 hpf samples induced a significant down regulation of this genes. These genes have then been considered as a training dataset. Indeed, they displayed a robust decrease in expression levels when we compared both dnFGFR to control and dnFGFR to EtsVp16, at 8hpf and 12 hpf, as reported in the heat map.

In order to enrich the training dataset, I took the Table S2 (Haeussler et al., 2010) as reference, where the annotation of expression pattern in CNS for 258 genes at early/middle tailbud stages is reported. Among them I specifically selected a set of genes expressed in dorsal anterior sensory vesicle of the *Ciona* CNS, a region comprising pigment cell lineage territories; I next checked which of them were differentially expressed in the gene lists obtained from microarray data analysis and selected *Ci-Prickle1*, *Ci-NetA*, *Ci-Noto2*, *Ci-FoxB*, *Ci-Aebp2*, *Ci-Bcmo1*, *Ci-Atpaf1*, *Ci-Mesk2* to add to the training dataset ; even if their spatial expression is widely extended in the Central Nervous System (CNS), all of them are expressed also in the territories where the PCPs are located. As known PCP markers have a differential expression profile when FGF signalling is perturbed, also this set of genes show a robust differential expression pattern, either positively or negatively, both at 8hpf and 12 hpf stages (Fig. 3.15). These data indicate that microarray dataset provides an accurate representation of genes expressed in PCPs territories.
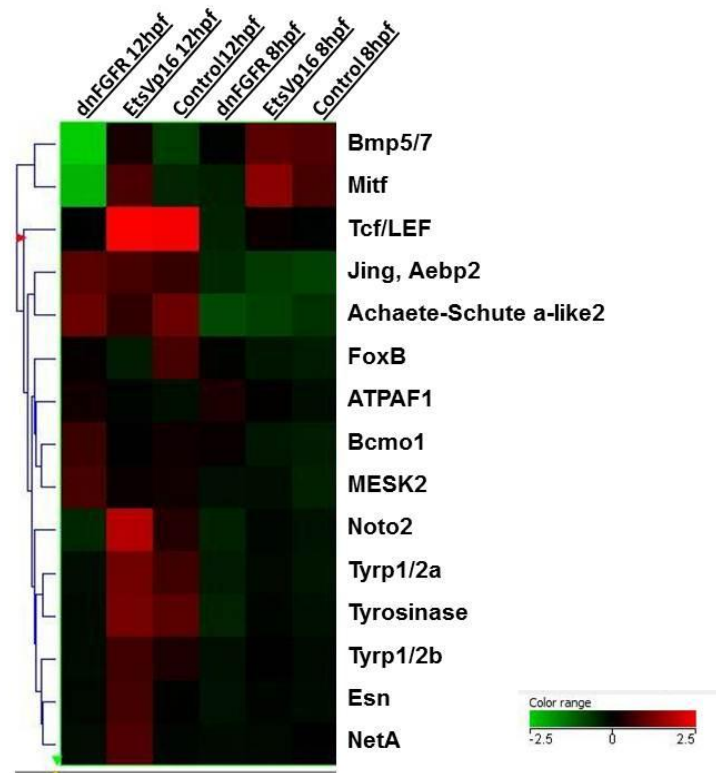
Fig. 3.15 Hierarchical clustering and heat map of the training set data expression values. Clustering was performed on all six transcript profiled samples, referred to the three conditions (control, dnFGFR, eTSvP16) at the two developmental stages (8 and 12 hpf). Expression levels are colour coded with shades of red, green and black corresponding to an increase, decrease and no change in gene expression, respectively, as represented by the colour range.

3.7 *Selection of candidate genes for biological validations*

Candidate genes regulated by FGF/MAPK/ETS signalling, either positively and negatively, have been selected from the gene list displaying significant (p-value < 0.05) and a robust differential expression (fold change > 1.5) when comparing the two conditions, dnFGFR and EtsVp16, to control at 8 and12 hpf (Appendix V, Table1-3); in particular t-test analysis on the microarray dataset identified 958 probe sets displaying significant changes when comparing dnFGR to control samples at 12hpf; of these, ~300 genes are annotated on *C. intestinalis* genome. When comparing EtsVp16 and control samples at 12 hpf, 905 transcripts are differentially expressed; of these, ~300 genes are annotated in *C. intestinalis* genome.

The statistical analysis performed on 8hpf samples revealed that lower number of transcripts are differentially expressed in comparison to 12 hpf samples and I mainly focus on the comparison of dnFGFR and control samples that evidence 144 transcripts with a significant changes; among these, 60 genes are annotated in *C. intestinalis* genome, while comparing EtsVp16 to control samples 260 probe sets have significant changes but none shows a fold change higher 1.5; for this reason, I considered just the significant differentially expressed transcripts identified from the comparison dnFGFR and control samples. Among the transcripts with significant differentially expression that are annotated in *C. intestinalis* genome (Appendix V, Table1-3), I firstly selected all genes encoding for transcription factors and cell signalling molecules with the final aim to find new FGF/MAPK/Ets target genes involved in the cascade or in the interactive communications with other signalling. In ascidians, cell–cell interaction plays a pivotal role in the specification at least of embryonic cells that give rise to CNS, notochord and mesenchyme. I also considered genes that have been identified from GO analysis (Appendix III), paying particular attention to the GO classes referred to pigmentation, melanocytes differentiation Myosin V binding transport, negative regulation of Wnt receptor signaling pathway; moreover I selected transcripts from the neighbors analysis on transcriptional regulatory network we created (Appendix IV, Table1-5) to identify new candidates that are co-expressed with melanogenic enzymes and could be implicated in pigment cell formation; I checked one-by-one on ANISEED database (Ascidan Network for in situ Expression and Embryological Data) (Tassy et al., 2010) if the expression pattern is reported in *Ciona* and I selected, for further studies, all of them that have not been characterized yet. In order to start the biological validation of microarray dataset, I identified other candidate target genes using criteria based on biological function of genes already characterized in other organisms; I mainly focused attention on genes associated to regulation of development of

Central Nervous System (CNS), involved in formation of anterior CNS or in pigmentation processes or in Wnt and BMP signaling activation. I also considered genes that in other organism modulate and integrate Bmp and canonical Wnt signalling, other genes have been chosen because involved in FGFR, MAPK or Ets regulatory feedback loops that are likely to modulate signaling dynamics following initial specification; I also focused on genes involved in G-protein coupled receptor protein signaling pathway such that are important for the formation of retinal pigmented epithelium; moreover I considered genes that in other organisms are responsible for melanosomes trafficking, lysosome formation, cytoskeleton function such as endo- and exo-cytosis, gap junction formation. Although there are few evidences of melanosomes organization in *Ciona* these genes may serve also in this organism to prime melanosomal logistics including biogenesis and transport. Among all the genes selected, 80% had no expression data in ANISEED expression database; in particular from the list of ~300 probe sets differentially expressed when comparing dnFGFR to control at 12 hpf, I selected 45 genes (Appendix IV-Table 1), form the ~144 transcripts with differential expression when comparing of dnFGFR to control at 8 hpf, I identified 13 genes (Appendix IV-Table2) and from the ~300 transcripts with significant differential changes when comparing EtsVp16 to control at 12 hpf, 34 genes (Appendix IV-Table3). The comparison of the 3 selected gene sub-groups revealed that just one gene is present when comparing the 3 groups together while 19 genes are differentially expressed in both conditions, dnFGFR and EtsVp16, compared to control samples at 12 hpf; moreover 9 genes result commonly present when dnFGFR is compared to Control both at 8 and 12 hpf (Fig. 3.16).
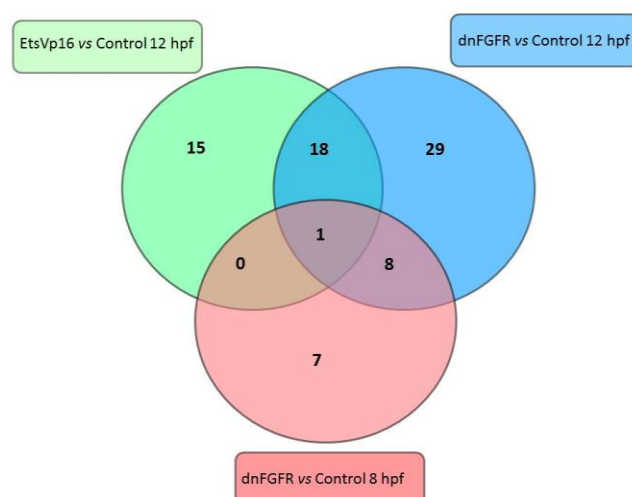
Fig. 3.16 Identification of 78 genes for further biological validation. Venn Diagram displaying the distribution of significant differentially genes among the three comparison: EtsVp16 to control at 12hpf (green), dnFGFR to control at 12hpf (blue)and dnFGFR to control at 8 hpf (pink).

3.7.1 Validation of the FACS/microarray approach by in situ hybridization

The experimental approach to validate microarray data has been using whole mount in situ hybridization (WMISH) of candidate genes, which provide independent information about spatial and temporal gene expression. To mirror microarray experiment, in-situ expression assays have been performed at neurula (8hpf) and tailbud stage (12hpf) *Ciona* embryos. In addition, double in situ assay have been used to reveal the co-localization of a new candidate with a known marker gene of a specific territory (i.e. *Ci-Tyrp1/2a* for PCPs or *Ci-Six3/6* for anterior sensory vesicle territories).

3.8 *FGF signaling is important to establish the fate of anterior neural plate*

When FGF signaling is perturbed in PCPs, a substantial subset of probes display an increased expression levels in dnFGFR samples and a decreased expression pattern in EtsVp16 in comparison to control samples. Among these genes, *Ci-Six3/6* is the only ortholog of vertebrate *Six3* and *Six6* (Wada et al., 2003), is expressed in the presumptive antero-ventral sensory vesicle and the anterior neural boundary at early tailbud stage; expression became progressively restricted anteriorly to the sensory vesicle and was excluded from stomodeum in the hatching larval stage; at this stage, *Ci-Six3/6* is weakly expressed in the vicinity of ocellus and in the posterior sensory vesicle (Moret et al., 2005). In literature it's reported that the maintenance and refinement of anterior neural fates requires that Wnt signaling is transcriptionally repressed in the anterior neuroectoderm, and *Six3*, a repressor factor of Wnt receptor signaling, is a key player during this process (Lagutin, 2003). In *C. intestinalis* both FGF and the canonical Wnt pathways are involved in pigment cell induction and differentiation; these two signaling pathways cross-talk by the direct transcriptional regulation of FGF cascade, via the MAPK/Ets, on the Wnt downstream effector gene *Ci-Tcf* (Squarzoni et al., 2011). Considering *Ci-Six3/6* spatial expression pattern and its robust increased values when FGF cascade is perturbed, I consider the hypothesis that FGF activity specifies different anterior-to-posterior fates within the neural plate and it may be required to inhibit *Ci-Six3/6* expression in the posterior territories where pigment cell precursors are located.

I firstly demonstrated that *Ci-Six3/6* transcripts are localized in a territory that is adjacent but not overlapping with *Ci-Tyrp1/2a* expression region from neurula stage (Fig. 3.17) till tailbud stage, by double WMISH.
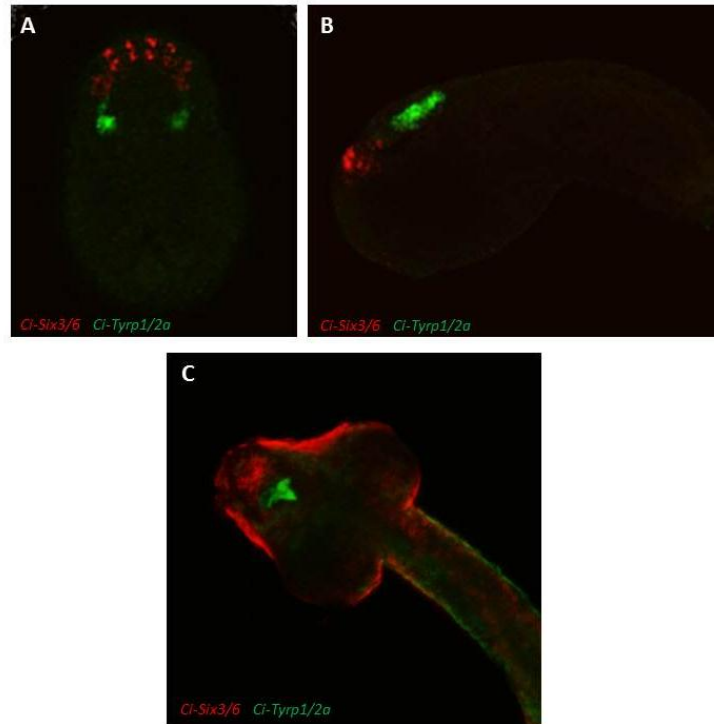
Fig. 3.17 Double WMISH with Fluorescent Tyramide staining for *Ci-Six3/6* (red) and *Ci-Tyrp1/2a* (green) on neurla, tailbud and young larva stage of *Ciona* wild type embryos (A-C); at neurula stage (A), *Ci-Six3/6* is expressed in the anterior region of the neural plate, adjacently but not in overlapping with PCPs, that are marked by *Ci-Tyrp1/2a* expression; this expression pattern is maintained at tailbud stage (B, lateral view) and also at young larval stage (C, dorsal view), where *Ci-Six3/6* is expressed in the anterior CNS.

Microarray data showed that *Ci-Six3/6* has a robust increase when comparing dnFGFR and control samples, in one developmental stage (12hpf). To understand where *Ci-Six3/6* transcripts are localized in embryos where FGF signaling was perturbed, I performed a double in situ expression assay for *Ci-Six3/6* and *Ci-Tyrp1/2a*; this experiment revealed that inhibition of FGF signaling caused ectopic *Ci-Six3/6* expression in a posterior region that includes the precursors of *Ci-Tyrp1/2a*-positive PCPs. Conversely, it result slightly inhibited in the EtsVp16 condition (Fig 3.18 A-C').
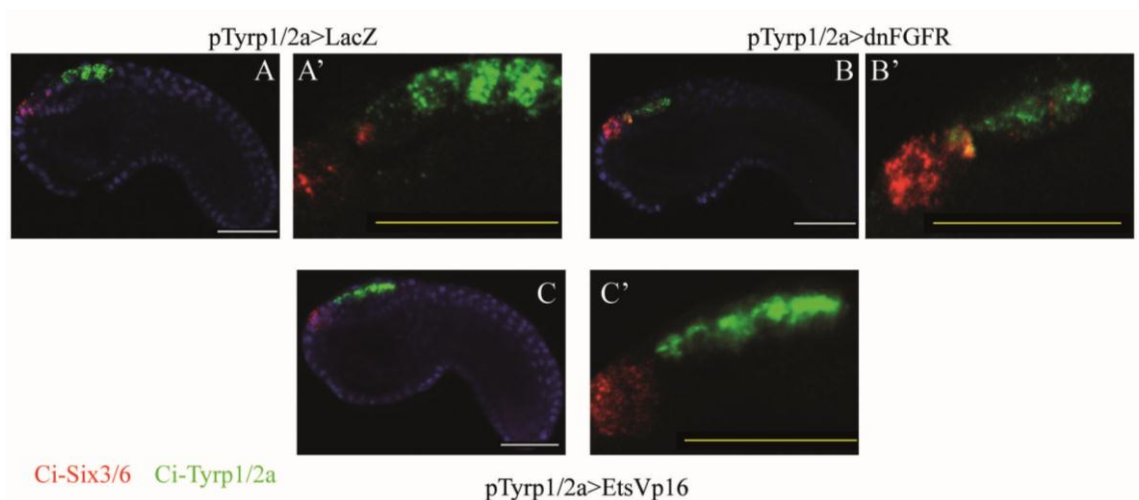
Fig.3.18 Double WMISH with Fluorescent Tyramide staining for *Ci-Six3/6* (red) and *Ci-Tyrp1/2a* (green) on *Ciona* tailbud embryos electroporated with plasmids pTyrp1/2a>Lacz (A),

pTyrp1/2a>dnFGFR (B) and pTyrp1/2a>EtsVp16 (C). (A) *Ci-Six3/6* is expressed in the anterior region of the presumptive sensory vesicle and it does not overlap with *Ci-Tyrp1/2a* domain; tailbud embryos expressing pTyrp1/2a>dnFGFR transgene shows an ectopical signal in a posterior region that include the anterior precursors of PCPs as marked by the co-expression with *Ci-Tyrp 1/2a.* (C) embryos electroporated with pTyrp1/2a>EtsVp16 have *Ci-Tyrp1/2a* transcripts endogenously present in PCPs, but show also a more anteriorly ectopically expression; in these embryos, *Ci-Six3/6* is expressed in the anterior part of CNS although the expression in the most posterior domain results very weak. DAPI staining for nuclei visualization.

These observations support the notion that when dnFGFR is expressed in the PCPs, genes that are normally inhibited by FGF signalling became ectopically activated in *Ci-Tyrp12a* positive cells.

In order to find other genes that show expression profiles similar to *Ci-Six3/6*, I identified all genes displaying expression changes similar to *Ci-Six3/6* among the list of 75 genes selected from statistical analysis. Among the "regulated-positive" candidates that show an increased expression upon dnFGFR expression, I focused on few genes that are implicated in CNS development and with a regulative roles in nervous system formation, such as *Ci-Pax3/7, Ci-Rgr/Nut2, Ci-Neurogenin, Ci-FoxHa, Ci-EBF/COE, Ci-SLC7A9, Ci-Macho1* (Fig.3.19).
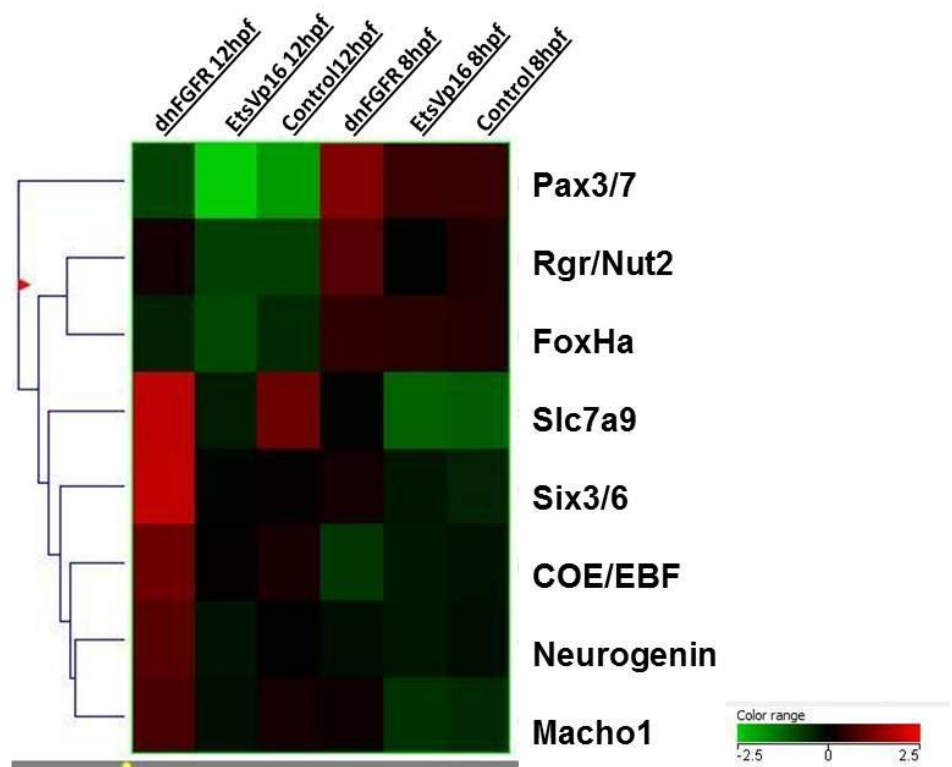
Fig. 3.19 Hierarchical clustering and heat map for expression values of genes Up-regulated when FGF signaling is blocked. Clustering was performed on all six transcript profiled samples, referred to the three conditions (control, dnFGFR, EtsVpP16) at the two developmental stages (8 and 12 hpf). Expression levels are colour coded with shades of red, green and black corresponding to an increase, decrease and no change in gene expression, respectively, as represented by the colour range.

These genes are already known to be involved in formation of the nervous system in Vertebrate and have a conserved role in *Ciona* too; furthermore their spatial and temporal expression profiles are available on ANISEED website, except for two genes: *Ci-Rgr/Nut2* and *Ci-SLC7A9*.

Pax3/Pax7: in *Ciona* there is one homolog of human PAX3 and PAX7, *CiPax3*/7, that starts to be expressed at neurula stage in three small regions along the ridges of the neural plate and in the ectoderm cells adjacent to the neural plate. At tailbud stage, the expression is limited to anterior and posterior bands in the brain and to dorsal cells in the posterior neural tube and posterior epidermis; at the late tailbud stage, the expression persists in the anterior sensory vesicle and two narrow stripes in the posterior brain, while posteriorly transcripts are localized in the dorsal neural tube and epidermis (Mazet et al., 2003).

Neurogenin*: Neurogenin* encodes a basic helix-loop-helix (bHLH) transcription factor and in *Ciona* its expression is restricted to the lateral-most A-line neural blastomere, A8.16, at the early gastrula stage (Imai, 2004); at early tailbud *Ci-neurogenin* is expressed in the anterior region of the presumptive brain. No expression was detected outside of the central nervous system at any embryonic or larval stage.

EBF/COE*: COE* encodes a transcription factor of the Collier/Olf/EBF family (Satou et al., 2003b) and *Ci-COE* transcripts are present in row II/column 4 of the neural plate stage (A9.32) (Imai et al., 2006). Expression of *Ci-COE* was detected in the early tailbud embryos in a bilateral pair of cells in the rostral trunk epidermis and in a second more posteriorly placed bilateral pair of cells in the neural tube; at the middle and late tailbud stage the expression is detected in the anterior central nervous system (Mazet et al., 2005).

FoxH1: the winged helix transcription factor FoxH1/Fast1/Schmalspur is required for the specification of the hypothalamic fate in diencephalic cells by transducing Nodal signals emitted from the mesendoderm; *Ci-FoxHa* is expressed in various regions at late neurula and early tailbud stages: caudal epidermis and mesoderm, anterior nerve cord, and a transverse circular domain at the rostral epidermal tip of the embryo. In the neural domain, *Ci-FoxHa* displays a broad expression in the sensory vesicle encompassing most of the lateral and anterior ventral part (Moret et al., 2005).

Macho-1: in *C.intestinalis Macho-1* is expressed in cells of the central nervous system after neurulation: a strong signal was first seen in posterior cells of the nerve cord, and a weak signal in the anterior sensory vesicle at neurula stage; at tailbud stage *Ci-Macho-1* is expressed, in the anterior sensory vesicle, visceral ganglion and nerve cord precursors (Satou et al., 2002).

Among the group of genes I found with a similar expression variation to *Ci-Six3/6* when FGF signalling is perturbed, *Ci-Rgr/Nut2* and *Ci-SLC7A9* are the only two genes that have not been characterized during *Ciona* embryogenesis.

Ci-Rgr/Nut2 is the homologue of human RPE-retinal G protein-coupled receptor; in *Ciona* it encodes a G-protein coupled receptor and has a neural tube-specific expression; *in situ* assay revealed that *Ci-Rgr/Nut2* is expressed in neural plate cells and later during neurulation anteriorly in the CNS and along the entire length of the neural tube.

In *C. intestinalis*, *slc7a9* is the ortholog of vertebrate Large Neutral amino acids transporters Small subunit L type amino acid transporter; interestingly *Ci-slc7a9* has a very restricted neural expression; by WMISH I detected signals localized in an anterior domain of the CNS at tailbud stage. In order to demonstrate that this subset of genes are expressed in the anterior CNS region where also *Ci-Six3/6* transcripts are present, I performed double WMISH, using fluorescently labeled *Ci-Six3/6* as a marker gene for the anterior sensory vesicle territories (Fig. 3.20); even if they show an expression pattern in different territories of the tailbud stage embryo, I detected that all these genes are expressed in a restricted part of the anterior neural tube, that is partially overlapping with the territory where *Ci-Six3/6* transcripts are expressed (Fig. 3.20 A'-G').
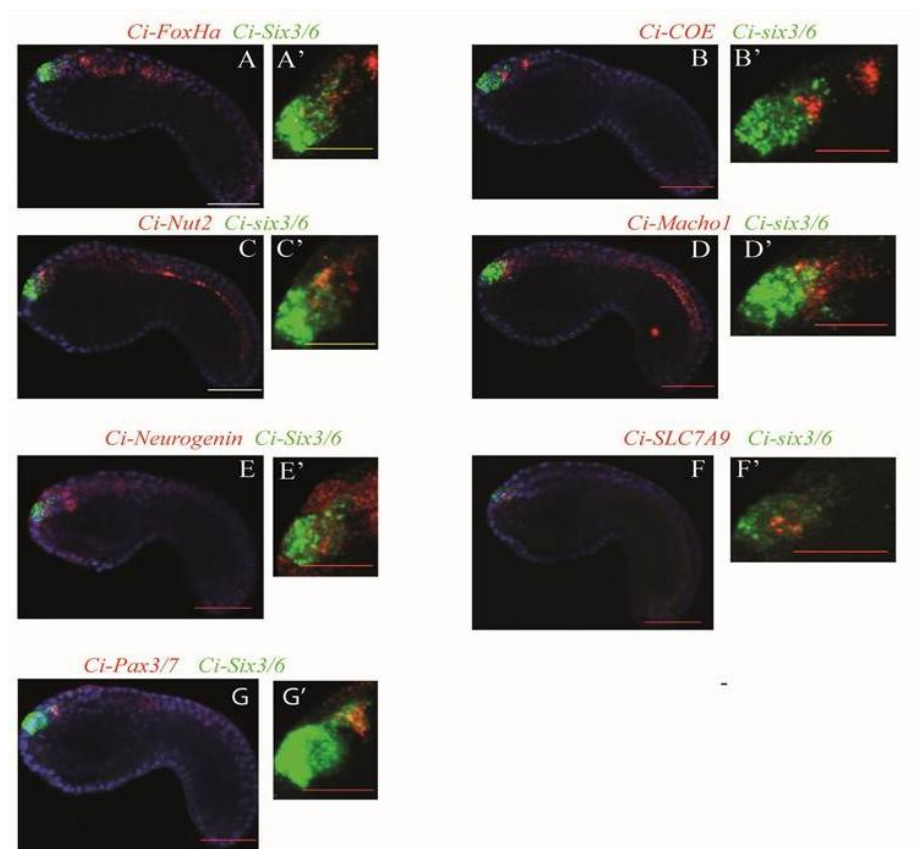
Fig. 3.20 Double WMISH with Fluorescent Tyramide staining for a select set of genes upregulated when comparing dnFGFR and control samples (A-G). *Ci-FoxHa* (A), *Ci-COE (B)*, *Ci-Nut2*(C), *Ci-Macho1*(D), *Ci-Neurogenin* (E), *Ci-SLC7A9* (F), *Ci-Pax3/7* (G) expression domains (red) are compared to *Ci-Six3/6* expression (green), considered as a marker gene for the anterior sensory vesiscle; zoom on the co-

localization region between *Ci-Six3/6* and the sub-set of genes in the anterior CNS (B''-G''). DAPI staining for nuclei visualization.

These observations support the notion that FGF signalling could be have a role in the inhibition of the anterior CNS specification in the pigment cell precursors.

According to the lineage studies, at late gastrula stage, *Ci-Six3/6* is expressed in the neural plate fourth row of cells, including in a9.50 blastomers. These cells derives from the division of a8.25 blastomeres and are the sister cells of the PCPs, the a9.49s located on neural plate III row; the a9.49s are in close proximity to two sources of FGF diffusible signaling molecules, Ci-FGF8/17/18 and Ci-FGF9/16/20, while *Ci-Six3/6+* positive cells are localized anteriorly than PCPs, resulted in an upper row of the neural plate cells and more distant from FGF sources. Furthermore this two gene expression patterns persist also at tailbud stage, where PCPs are located lined in the posterior neural tube while *Ci-Six3/6* is expressed more anteriorly in the presumptive antero-ventral sensory vesicle (Fig. 3.21).
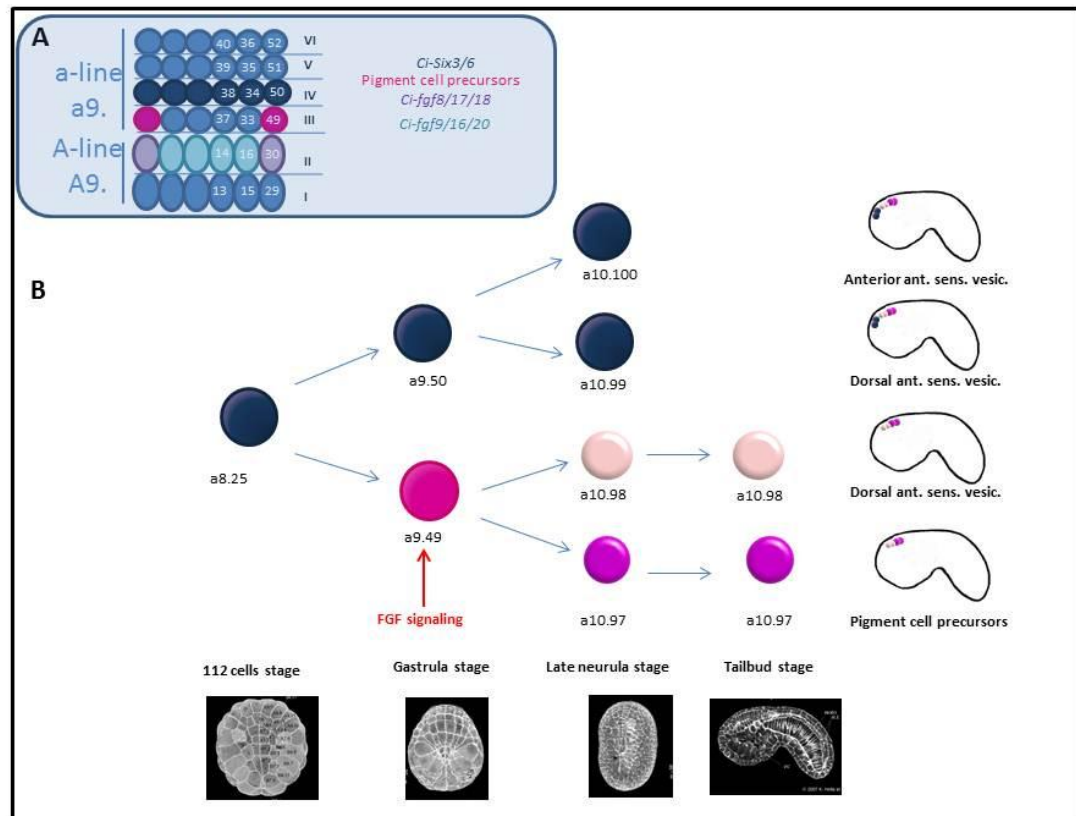
Fig. 3.21 Cell-lineage scheme of PCPs and anterior CNS formation (A,B). (A) At late gastrula stage, A-line nerve cord cells of the neurula plate express in the II row cells, *Ci-fgf9/16/20* (light green) and *Ci-fgf8/17/18* (violet) and *Ci-Six3/6* in all the IV row (blue). (B) PCPs arise from a8.25 cell pair division up to pigmented cell at larval stage. See discussion for a step-by-step description of the model. In blue are shown precursors of Anterior CNS, that arise from a9.50 cell division; in light pink are shown the a10.98s and in pink the a10.97s.

Indeed when dnFGFR is expressed at tailbud stage, *Ci-Six3/6* is ectopically expressed in a posterior region of the CNS, as compared to the control, where PCPs are present. All these

considerations support the notion that FGF signalling is required to induce the a9.49s as PCPs and that in absence of FGF molecules the a9.49s changes their fate, expressing more anterior CNS genes (Fig. 3.22). Thus it appears that FGF signaling cascade is broadly required for proper neural system differentiation and anterior-poster patterning, in particular to establish the proper identity of cells in row III and IV.
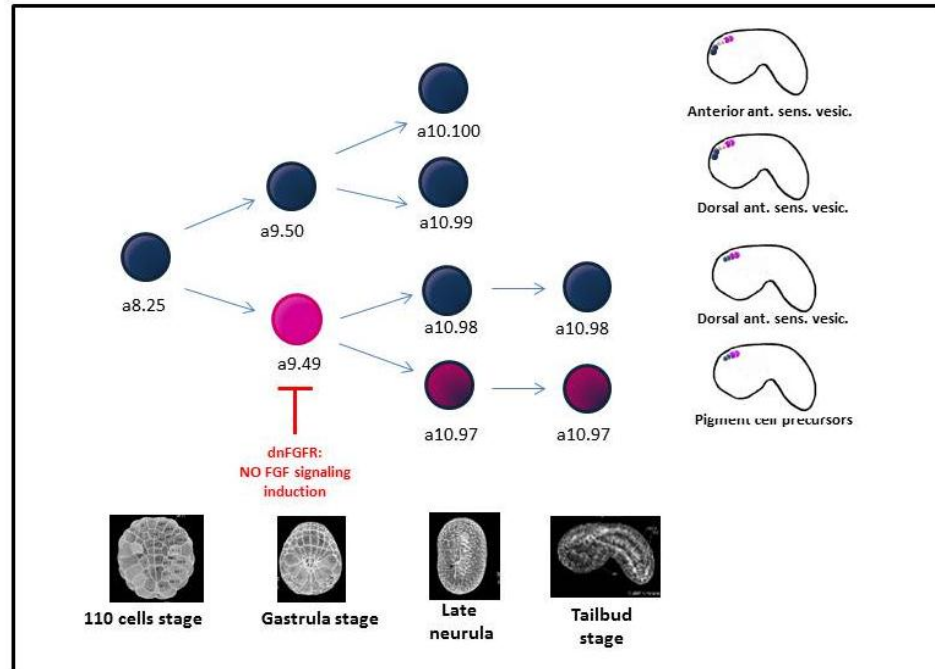
Fig. 3.22 Cell-lineage scheme of PCPs and anterior CNS formation when FGF signaling is blocked. Without the influence of FGF signaling, the a9.49 cell pairs are not fate restricted as PCPs; they change their fate, expressing more anterior CNS genes. In blue are shown precursors of Anterior CNS, that arise from a9.50 cell division; the blue-violet cells represent the a10.98s and the a10.97s, that behave like cells present anteriorly in the brain

3.9 *Biological validation of new candidate genes expressed in PCPs*

To find novel and uncharacterized genes involved in pigment cell formation, I prepare a set of cDNA clones to be used for expression profile and functional analyses. Among the gene list of new candidates selected after statistical analysis, I identified 50 candidates that could be good candidates for their biological functions, as previously described (Par. 3.7). I firstly checked on Aniseed database if they have been characterized and if the expression pattern was reported; 90% of them had no expression data in ANISEED expression database. In order to study their spatial and temporal expression, I prepared an antisense riboprobes complementary to the transcript; to this aim, I selected for each genes the correspondent location of cDNA clones in Nori Satoh *C. intestinalis* gene collection library, available in the laboratory. I picked the right clones from the cDNA library and let them grew at 37°C for a Mini-preparation DNA. Each clone selected has been confirmed to be corresponded to the gene chosen by DNA sequencing; the cDNAs have been digested

and retro transcribed to obtain labeled anti-sense riboprobes to use for in situ assay. Once obtained riboprobes, I screened candidate genes by in situ hybridization analysis on all 50 genes selected in order to detect signals from the territory where they are expressed; not all ISH experiments gave me good results; for some of them I was unable to detect signal at all, although they have a good detectable expression levels from microarray dataset. Among the 50 genes screened by WMISH, I observed a good expression assays for 30 genes; overall, 75% of which displayed expression in the sensory vesiscle precursors where PCPs are located. This result indicate that microarray dataset obtained provides an accurate representation on genes expressed in the central nervous system, in particular in PCPs territory. In this sub set of 30, 8 genes are specifically expressed in PCPs, *Ci-Lrp4, Ci-Doc2a, Ci-Bzrap1, Ci-Casz1, Ci-Mad/mnt, ,Ci-SLC45A2, Ci-Piwi, Ci-Rab38/32* and I considered them as new *Ciona* candidates in pigment cell lineage. For the first seven, no expression pattern is reported on ANISEED. Just for *Ci-Piwi* the expression profile was already known; Satou and others described Ci-PIWI as Drosophila dorsal switch protein 1 homolog, expressed in *Ciona* only in CNS, in a subset of brain cells (Satou et al., 2003a). Although it has been characterized in *Ciona* embryos, no evidences are described about its expression in PCPs or a putative role in pigment cell formation.

Low density lipoprotein-receptor: *Lrp4* is a multifunctional member of the low density lipoprotein-receptor gene family and a modulator of extracellular cell signaling pathways in development. For example, Lrp4 binds Wise, a secreted Wnt modulator and BMP antagonist signalling (Ohazama et al., 2008); physiological functions for Lrp4 include the endocytosis of a large number of macromolecules, including lipoproteins, proteases and protease inhibitors, as well as functions as direct signal transducers or modulators of several fundamental signal transduction pathways, including BMP and canonical Wnt (Choi et al., 2009).

In *Ciona* tailbud stage embryo *Ci-Lrp4* is expressed in all the four PCPs lined at the dorsal midline of the tailbud embryos central nervous system; a strong signal is also present posteriorly in the nervous system and in the tail muscles (Fig. 3.23 E-E'').

Doc/Rabphillin: Doc2/raphillin belong to the family synaptotagmin-related molecules; Rabphillin are a small GTPase Rab3A/Rab27A effector proteins family that ,together with Doc2α/β and Synaptotagmin, has an important role in docking and/or priming steps of dense-core vesicle exocytosis (Sato et al., 2010). In *Ciona* tailbud embryos, *Ci-Doc2* transcripts are localized just in two out the four cells that correspond to PCPs, although the signal detected has very low intensity (Fig. 3.23 D-D'').

Benzodiazepine receptor-associated protein: BZRAP is a peripheral-type benzodiazepine receptor-associated protein1, belonging to RIM-binding proteins (RIM-

BPs) that were identified as binding partners of the presynaptic active zone proteins RIMs as well as for voltage-gated Ca^{2+}-channels. They form a functional link between the synaptic-vesicle fusion apparatus and Ca^{2+}-channels (Mittelstaedt and Schoch, 2007). RIM proteins are an integral part of the cytomatrix at the presynaptic active zone and are coupled to synaptic vesicles via their interaction with the small GTPase Rab3; Rab3 is a neuronal GTP-binding protein that regulates fusion of synaptic vesicles and is essential for long-term potentiation of hippocampal mossy fibre synapses (Wang et al., 1997). In *Ciona*, *BZRAP* is restrictively expressed in all four PCPs lineage at tailbud stage; the intensity of the signal detected is very strong and *Ci-BZRAP* transcripts have a localization very close to the cell nucleus (Fig. 3.23 B-B'') .

Hunchback: the gene *hunchback* encodes a Cys2-His2 zinc finger transcription factor and plays a pivotal role in controlling pattern formation of the early Drosophila embryo (Wu et al., 2001). In *Ciona*, *hunchbacklike/casz1* has a wide expression in all the nervous system of the tailbud stage embryo; in particular *Ci-casz1* transcripts are localized all the four cells that correspond to PCPs (Fig. 3.23 G-G'').

Mnt/Mad: Mad proteins (Mad1, Mxi1, Mad3, Mad4, Mnt/Rox) are biochemical and biological antagonists of c-Myc oncoprotein. Mad-Max dimers repress the transcription of the same target genes activated by Myc-Max dimers. The varied activities of Max network proteins are manifested during proliferation and differentiation. Mad family proteins are generally induced during terminal differentiation and act to limit cell proliferation (Zhou and Hurlin, 2001). In *C. intestinalis* genome this transcription factor is annotated as *Ci-Mad/mnt/Rab24* and its transcripts are localized in PCPs at tailbud stage, although with a weak detected signals (Fig. 3.23 F-F'').

Solute carrier family 45, member 2: SLC45A2 (previously known as MATP, Membrane-associated transporter protein, or melanoma antigen AIM1) stands for solute carrier family 45, member 2 and is responsible in humans and corresponding mouse models for oculocutaneous albinism 4 (OCA4) , an autosomal-recessive condition characterized by reduced pigmentation of the hair, skin and eyes. SLC45A2 bears homology to sucrose–proton symporters in plants (Newton et al., 2001), so it might regulate osmoregulation within melanosomes; SLC45A2-deficient melanocytes show abnormalities in melanogenic enzyme trafficking and loss of mature melanosomes (Costin et al., 2003). In *Ciona* tailbud embryos, *Ci-slc45a2* is restrictedly expressed in all four blastomers that correspond to the PCPs (Fig.3.23 C-C'').

In order to demonstrate that all this genes are expressed in PCPs, I performed a double WMISH with fluorescently labeled transcripts of *Ci-Tyrp1/2a* that marks the four lined blastomers from pigment lineage at tailbud stage, the a10.98s and a10.97s (Fig. 3.23 A-

G''); to better visualize the co-localization area of selected genes and *Ci-Tyrp1/2a*, I employed confocal microscopy for the double fluorescent in situ assays.
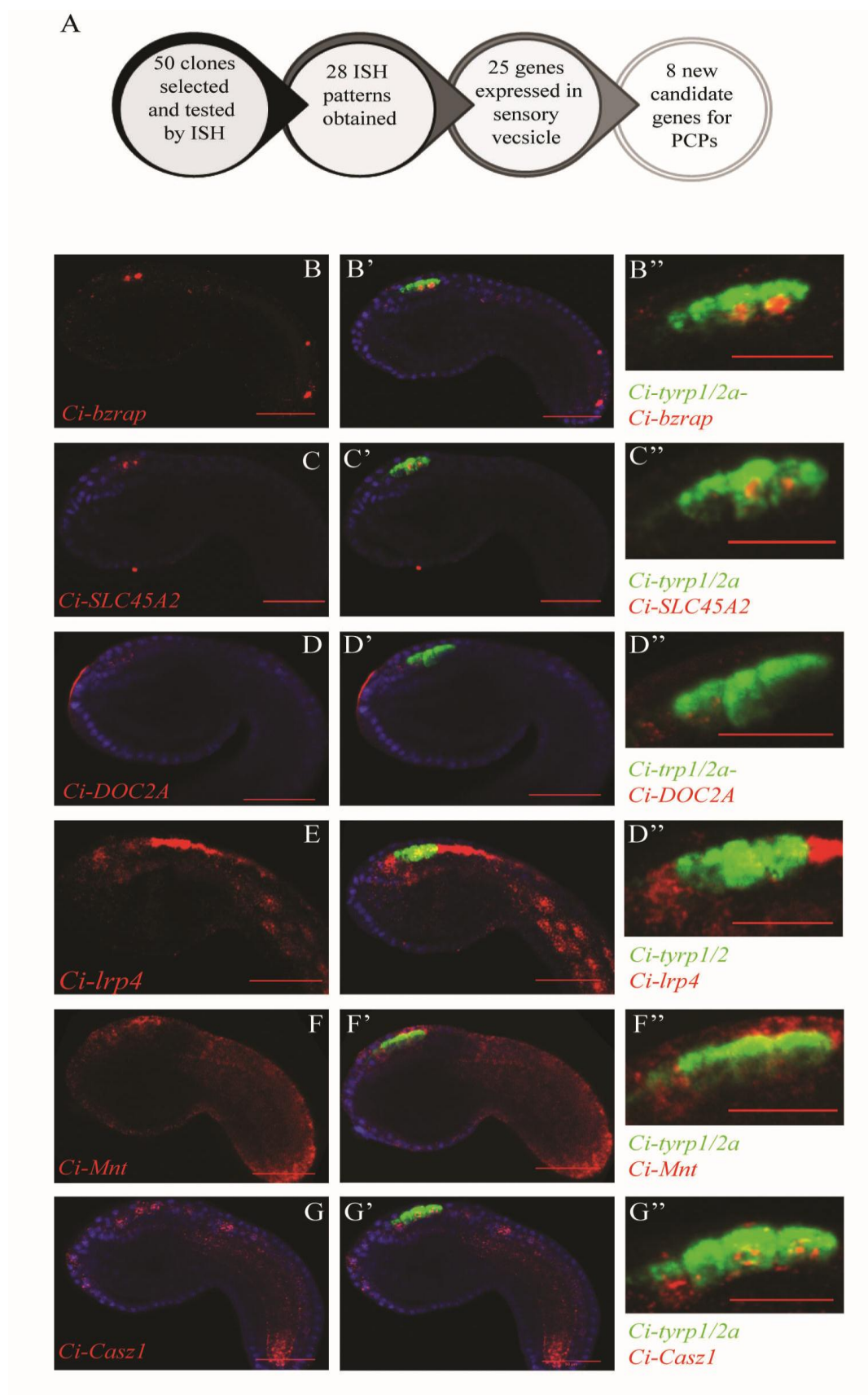
Fig. 3.23 New marker genes expressed in pigmented cell precursors (A-G): schematic representation of number of genes differential expressed when comparing dnFGFR and Control samples at12hpf and selected

for ISH experiments and their relative territories of expression (A); Expression of new *Ciona* candidate PCPs genes (B-G); double Fluorescent WMISH with new candidate genes (red) and *Ci-Tyrp1/2a* (green); (B) *Ci-bzrap*,(C), *Ci-slc45a2*; (D) *Ci-doc2a*; (E) *Ci-Lrp4*; (F) *Ci-Mnt*; (G) *Ci-Casz1*; Overlapping signals for green and red labeled transcripts (B'-G'); zoom on neural region where PCP s are located (B''-G''). DAPI staining for nuclei visualization.

3.10 *Ci-Rab38/32/rab-rp1/ltd* : a gene differentially expressed at 8hpf

By statistical analysis, I selected 958 differentially expressed transcripts when comparing dnFGFR vs Control at 12 hpf and 144 differentially expressed transcripts when comparing dnFGFR vs Control at 8 hpf demanding p-value < 0.05 and fold change > 1.5; the comparison of these two groups revealed that 44 differential expressed transcripts are commonly present between the them, 914 differential expressed changing specifically at 12 hpf while 100 transcripts specifically at 8 hpf (Fig. 3.24).
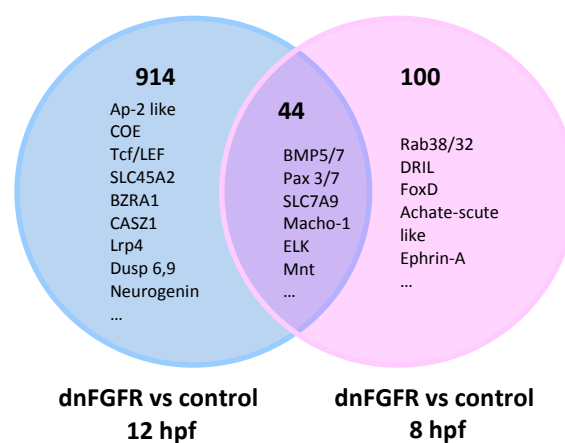
Fig.3.24 Identification of differentially expressed transcripts comparing dnFGFR to control lists at both 8 and 12 hpf. Venn Diagram displaying the distribution of significant differentially genes when comparing dnFGFR to control at 12hpf (blue) and dnFGFR to control at 8 hpf (pink).

Among the list of transcripts differentially expressed exclusively at 8 hpf, an interesting very good candidate that show a specific and restricted expression in PCPs is *Ci-Rab38/32/rab-rp1/ltd*; this gene is up regulated when comparing dnFGFR to control while it show a down regulation when EtsVp16 is compared to control at 8 hpf, as shown in the profile plot (Fig. 3.25).
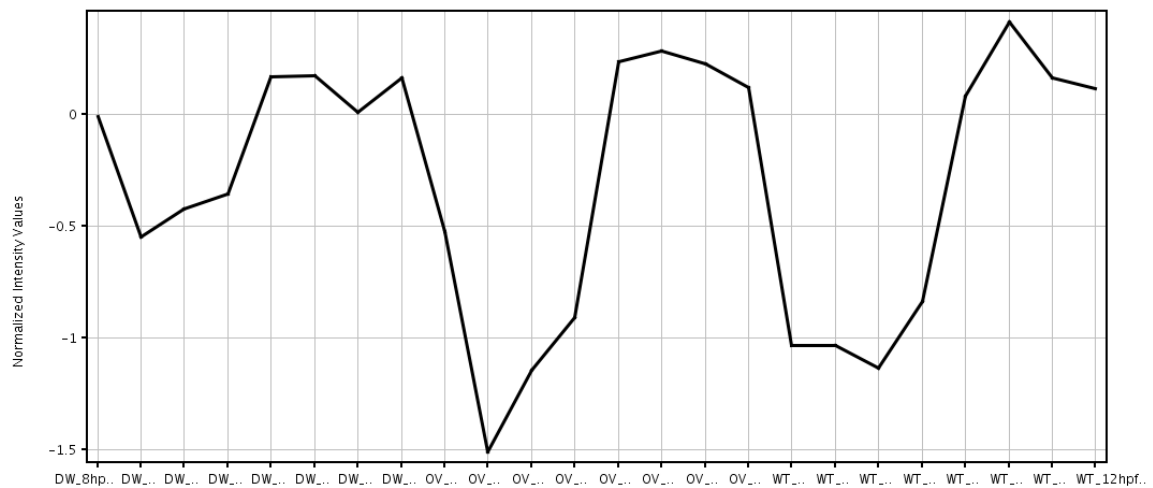
Fig. 3.25 Profile plot for *Ci-Rab38/32/rab-rp1/ltd*. On the *x*-axis, the expression values for the transcript associated to the gene *Ci-Rab38/32/rab-rp1/ltd* across all the samples, including the three conditions (DW=dnFGFR, OV=EtsVP16 and WT= Control) at 8 and 12 hpf; on the *y*-axis, the Normalized Intensity Values.

Rab32 and Rab38 are two highly homologous small GTPases belonging to the Rab family specifically expressed in melanocytes and localized on tyrosinase- and Tyrp1-containing vesicles and/or organelles around the trans-Golgi network (TGN) as well as on melanosomes (Wasmeier et al., 2006); it's noteworthy that the Rab-type small GTPases and their regulators are key factors involved in melanosomal logistic, a process that include melanosome biogenesis and transport in melanocytes (Ohbayashi and Fukuda, 2012).

3.10.1 Rab38/32/rab-rp1/ltd annotated in Ciona intestinalis genome

In order to identify how many genes encoding for Rab proteins are annotated in *C. intestinalis* genome, I carried out a bioinformatics research on Ensembl website (http://www.ensembl.org/index.html); the analysis revealed the presence of 39 genes encode for small GTPase belonging to Rab family; among them, a unique orthologous gene for mouse *Rab38* and *Rab32* is present, *Ci-Rab38/32/rab-rp1/ltd* that is contained in the Scaffold "HT000351.1:7,483-8,434" on the reverse strand annotated as transcript id "ENSCINT00000013519" (Fig. 3.26).
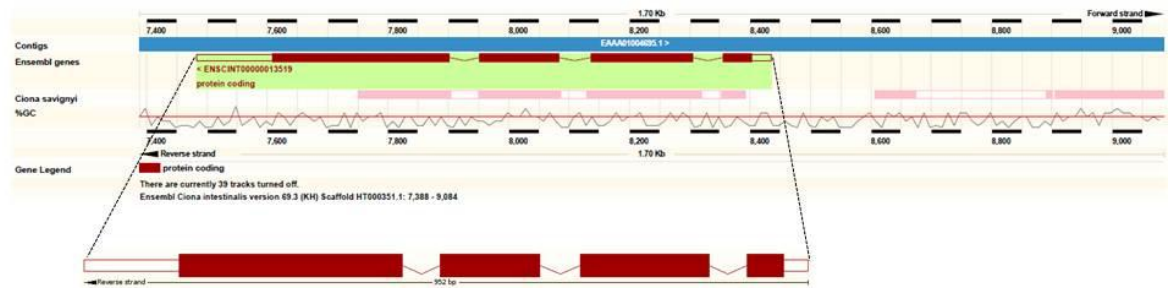
Fig. 3.26 *Rab38/32/rab-rp1/ltd* **gene location in the** *Ciona intestinalis* **genome.** Rab38/32/rab-rp1/ltd gene is located in the ciona4/Scaffold_ HT000351.1:7,483-8,434. The gene appears to be composed by 4 exons (in red) and 3 introns, extending for 952 bp (Ensembl gene prediction). Conserved transcript region with *C.savignyi* are evidenced as pink boxes; the black picks represent the%GC in the transcript. Adapted from Ensembl release 70 - January 2013 version of *C. intestinalis* genome assembly v 2.0 .

On the basis of the Ensembl gene prediction, the *Ci-Rab38/32/rab-rp1/ltd* gene appears formed by 4 exons and 3 introns, with a transcript length of 799 bp and a translation length of 213 amminoacids. A sequence of ~700 bp upstream to the ATG of the first *Ci-Rab38/32/rab-rp1/ltd* exon, separates it from the next upstream gene, that in Ensembl is annotate with the transcript id "ENSCINT00000013527".

3.10.2 *Ci-Rab38/32/rab-rp1/ltd expression pattern*

In order to study the spatial and temporal expression profile of *Ci-Rab38/32/rab-rp1/ltd*, I performed whole mount in situ hybridization experiments on *Ciona* embryos at different developmental stages observed by confocal microscopy (Fig. 3.27).

No signal was detected up to the late gastrula stage. The first expression of *Ci-Rab38/32/rab-rp1/ltd* has been observed at early neurula stage, in the a9.49 blastomere pair cells which corresponds to the PCPs (Fig. 3.27A). A clear and specific signal was then inherited in both a9.49 progeny (the a10.97 and a10.98 pairs) appearing much stronger in the posterior a10.97s, compared to the anterior a10.98 pairs, at late neurula stage, that correspond to the 8 hpf samples of microarrays. The expression persisted to the tailbud stage, in these four blastomers that line up along the dorsal midline of the developing neural tube (Fig. 3.23 B-E); although in contrast to the late neurula stage, the intensity of the signal is strongly detected in all the four precursors. The expression of *Ci-Rab38/32/rab-rp1/ltd* has not been detected at larval stage.
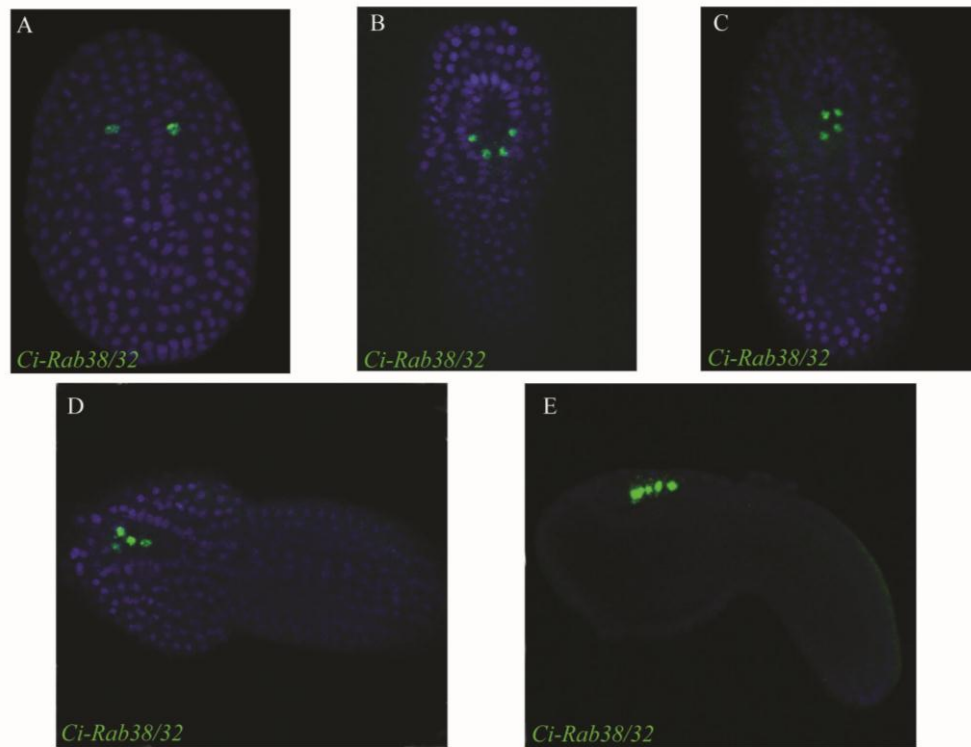
Fig. 3.27 *Ci-Rab38/32/rab-rp1/ltd* expression in pigment cell precursors (A-E) Expression of *Ci-Rab38/32/rab-rp1/ltd* (green) at neurula(A), initial tailbud (B), early tailbud (C), middle (D) and late tailbud (E); dorsal view (A-D) anterior at the top (A-C); lateral view, anterior to the left (E). DAPI staining for nuclei visualization.

In order to prove that *Ci-Rab38/32/rab-rp1/ltd* is expressed specifically in PCPs, I performed double fluorescent in situ hybridization using fluorescently labelled antisense riboprobes for *Ci-Tyrp1/2a* as PCP marker gene (Fig. 3.28); to mirror microarray I used embryos at neurula and tailbud stage and observed them with confocal microscopy. *Ci-Rab38/32/rab-rp1/ltd* is expressed in the a10.98s and a10.97 at neurula stage, with the posterior signals in a10.98s appearing stronger than the anterior in a10.97s, in comparison to *Ci-Tyrp1/2*a, that is strongly expressed in all the four PCP cells (Fig. 3.29 A-B). At tailbud stage, *Ci-Tyrp1/2a* and *Ci-Rab38/32/rab-rp1/ltd* transcripts are co-expressed in the same territories, in all the four cells lined along the neural tube that correspond to PCPs (Fig. 3.29 C-E''). Moreover *Ci-Tyrp1/2a* transcripts are widely detected in the cytoplasm of the PCPs, while *Ci-Rab38/32/rab-rp1/ltd* has a perinuclear localization.
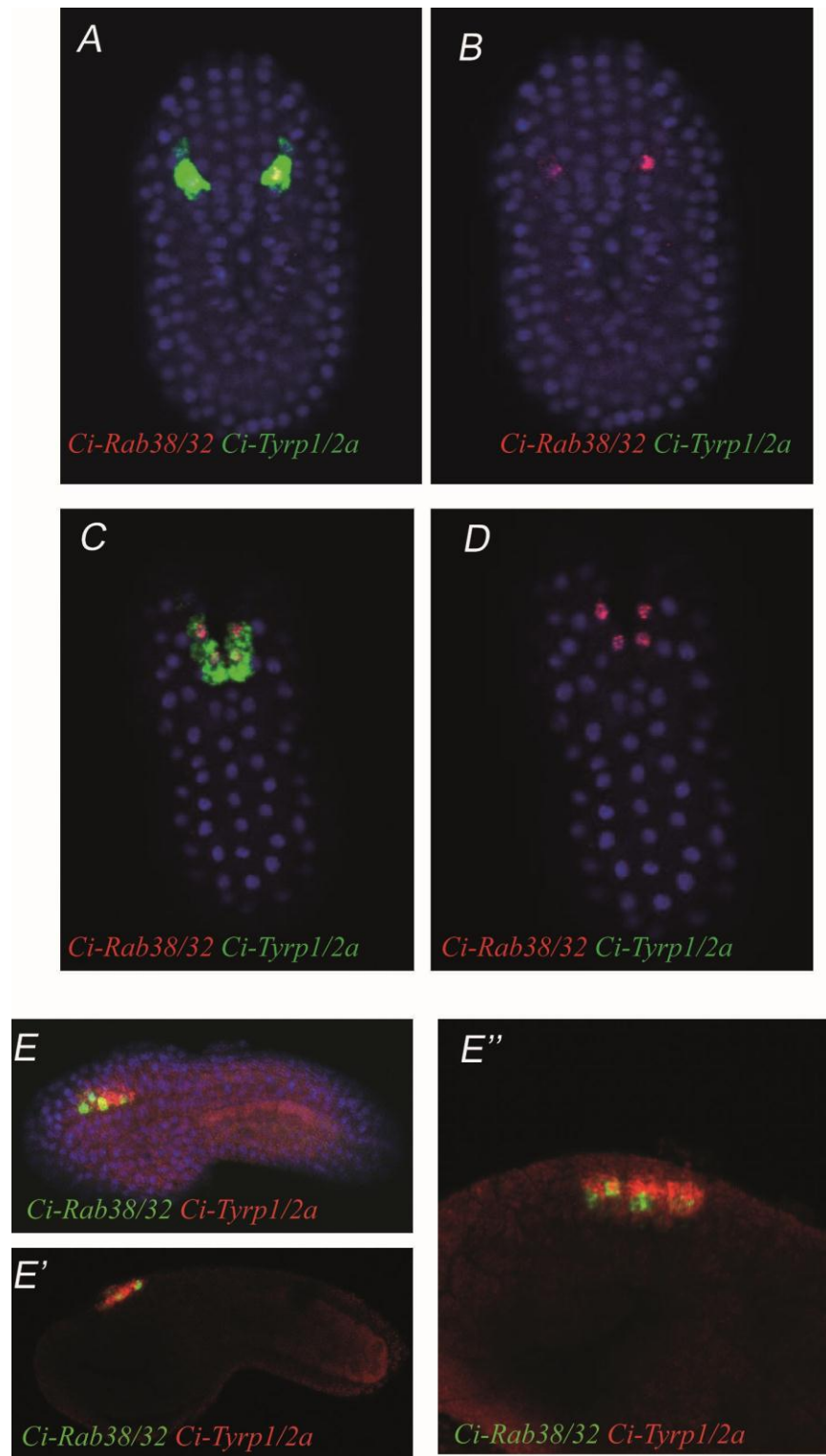
Fig.3.28 Double WMISH with Fluorescent Tyramide staining for *Ci-Rab38/32/rab-rp1/ltd* and *Ci-Tyrp1/2a* from neurula to tailbud stage of *Ciona* embryos (A-E); at neurula stage (A,B), *Ci-Rab38/32/rab-rp1/ltd* (red) is expressed in the posterior PCPs, the a10.98s, *Ci-Tyrp1/2a* (green) expression is detected in all four PCP cells; at early tailbud stage (C,D), *Ci-Rab38/32/rab-rp1/ltd* (red) is expressed in all four PCPs, colocalizing with *Ci-Tyrp1/2a* transcripts (red); at middle tailbud stage (E-E'', lateral view), *Ci-Rab38/32/rab-rp1/*ltd(green) and *Ci-Tyrp1/2a* (red) co-localize in the four PCPs, lined dorsally in the forming neural tube. Overlapping signals for green and red labeled transcripts (A,C,E,E',E''); zoom on neural region where PCP s are located (E''). DAPI staining for nuclei visualization.

As other monomeric Ras-like GTPases, Rab functions is regulated by a molecular switch cycling between two nucleotide-bound states, a GDP-bound inactive state and a GTP-

bound active state; in order to verify if *Ci-Rab38/32/rab-rp1/ltd* has conserved amminoacid sequence in switch regions I and II that are important for the switching cycle conformations, I performed a multiple sequence alignment using Clustal W2 algorithm from EBI website, (http://www.ebi.ac.uk/Tools/msa/clustalw2) selecting sequences for Rab32 and Rab38 protein from Human, Mouse, Xenopus and *Ciona* (Fig. 3.29).

In the two 'switch regions', Rab proteins possess a characteristic GTPase fold where the γ phosphate of GTP is contacted; an essential Mg^{2+} cofactor is required for high affinity nucleotide binding and hydrolysis. The Clustal W2 result confirms that the 'G motifs' in switch region I, involved in interactions with guanine nucleotides, (underlighted in yellow) is highly conserved in *Ciona* as well in the other Rab sequences; the "GxxxxGKT" motif provides phosphate contacts and supplies a Thr that contacts the γ phosphate and is coordinated by the Mg^{2+} ion in the active state (Vetter and Wittinghofer, 2001). Meanwhile the amminoacid sequence alignment evidences that the "DxxGQ" motif in switch II is highly conserved too; the switch II region corresponds to an interactive domain with the Rab effector providing an aspartate (D) that stabilizes the Mg^{2+} ion through interaction with a water ligand, a glycine (G) that contacts the γ phosphate, and a glutamine(Q) that functions as a catalytic residue for the intrinsic GTP hydrolytic reaction (Lee et al., 2009).

```
hRAB32   MAGGGAGDPGLGAAAAPAPETREHLFKVLVIGELGVGKTSIIKRYVHQLFSQHYRATIGV 60
mRAB32   MAGEGLGQQ--GASATAAPETREHLFKVLVIGELGVGKTSIIKRYVHQLFSQHYRATIGV 58
XRab32   MAGEGAGEE----QIPRAPERKEYLFKVLVIGELGVGKTSVIKRYVHQLFSQHYRATIGV 56
Ciona    ----------------MTEQRKEYLFKVLVIGELGVGKTSVIKRYVHQIFSQHYRATIGV 44
hRAB38   ---------------MQAPHKEHLYKLLVIGDLGVGKTSIIKRYVHQNFSSHYRATIGV 44
mRAB38   ---------------MQTPHKEHLYKLLVIGDLGVGKTSIIKRYVHQNFSSHYRATIGV 44
                         :*:*:*:****:*******:******* **.********
```

Fig.3.29 Multiple alignment for N-terminal Rab32 and Rab38 protein sequences. Amminoacid sequences of hRAB32 (*Homo sapiens* RAB32), hRAB38 (*Homo sapiens* RAB38), mRAB32 (*Mus musculus* RAB32), mRAB38 (*Mus musculus* RAB38), XRab32 (*Xenopus laevis* Rab32), Ciona (*Ciona intestinalis* Rab38/32/rab-rp1/ltd); sequence alignment was done by using the ClustlW2 algorithm from EBI website, (http://www.ebi.ac.uk/Tools/msa/clustalw2) using Default parameters. The yellow box represent the highly conserved 'G motifs' in switch region I.

Once identified a highly conservation in switch region I and II of *Ci-Rab38/32/rab-rp1/ltd* in comparison to other Rab32 and Rab38 amminoacid sequences, I created mutant Ci-Rab38/32/rab-rp1/ltd protein (hereafter referred to as Ci-Rab38/32 for simplicity) in order to evaluate the functional role of this protein during *Ciona* pigment cell formation. Sequence analysis of the Rab38 coding region from *chocolate* (*cht*) mutant mice revealed a G146T transversion in exon 1, results in a Gly to Val substitution in position 19 (G19V) (Loftus, 2002); the glycine 19 is located in "G motif" within the highly conserved phosphate/Mg^{2+} domain where GTP binding pocket is located. In order to test Ci-Rab38/32 protein function, I firstly created a dominant negative form (Rab38[G19V]) introducing a

single nucleotide substitution in codon 19 (GGT to GTT, as reported in *cht* mouse) that results in a Gly to Val substitution in *Ciona* switch I region; in this way, the Rab protein remains in a constitutive GDP-locked status and can not interact with its effector, resulting in an inactive form. To drive a directed expression of this dominant negative form, I used the *cis*-regulatory region of the melanogenic gene *Ci-Tyrosinase*, that is able to drive the expression of a reporter gene as mCherry from tailbud stage in PCPs to larval stage in pigmented sensory organs (Esposito et al., 2012); the construct has been tested by transgenesis *via* electroporation for the capability to direct pigment cell lineage-specific perturbation. Targeted expression of *pTyr>Rab38^{G19V}* constructs resulted in larvae showing defective pigmented cell melanization phenotypes in the sensory organs compared to wild type samples. In 39% of specimens, one pigmented cell appeared normally melanised (Fig. 3.30 A), whereas the other is absent or showing a partial melanization of the organ (Fig. 3.30 B); 13% of pertubated larvae have no pigmented sensory organs (Fig. 3.30 C), while 48 % of animals are apparently wild type, with the sensory organs well pigmented(Fig. 3.30D).
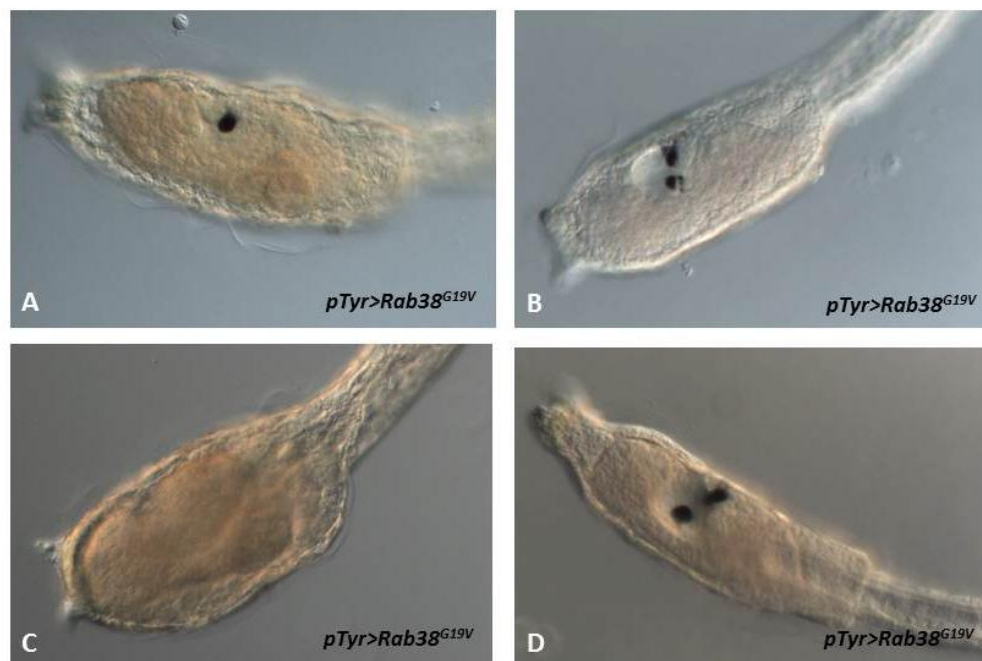
Fig. 3.30 pTyr>Rab38^{G19V} transgenic larvae (A-D). (A,B) 39% of the total screened larvae showed only one pigment cell in the sensory vesicle (A) or have both pigment cells but with a partial melanization of the organ (B); 13% of the total screened larvae showed complete absence of pigment cells(C); 48% of the specimens have both pigmented cells in the sensory organs appearing as wild type (D). Lateral view, anterior towards the left.

As statistics evidence, the Rab38^{G19V} mutant protein induced a faint alteration in *Ciona* pigmented sensory organ formation; this result is consistent if considering that the hypopigmented *cht* mice that are mutants characterized by a weakly diluted mouse coat color have been taken as reference for the mutation.

In literature is reporter that the most widely used mutations to alter nucleotide-binding properties involve asparagine substitution of the P-loop Thr (T-N mutant); this mutation in human Rab32 produces defecting proteins in GTP binding (Alto, 2002); Taking in notion this information, I created a double mutant protein, Rab38/32 (G19V+T23N) in order to have a stronger interference on Ci-Rab38/32 function; this mutant is characterized by two altered residues in "G motif" residues of the highly conserved switch I region: the first substation (G19V) is the same of *cht* mutant mice while the second one is a threonine to asparagine substitution in position 23 (T23N) of Ci-Rab38/32 protein. T-N substitution disrupts the Mg^{2+} binding site, resulting in greatly reduced affinity for guanine nucleotides. In vitro at low Mg^{2+} concentrations, the affinity for GTP is reduced 100-fold without affecting the affinity for GDP, suggesting that the mutants are 'GDP-locked' (Farnsworth and Feig, 1991; Nuoffer et al., 1994). The effects in vivo are more complicated, given relatively high cellular concentrations of free Mg^{2+} and GTP. In any case, the T-N mutant is expected to shift the equilibrium toward the GDP/nucleotide-free states and likely exert dominant negative effects by sequestration of endogenous GEFs due to low GTP affinity (Lee et al., 2009). In order to drive a directed expression of the double mutant in pigmented cell lineage, I used the promoter region *Ci-Tyrosinase* and test the construct obtained (pTyr> Rab38/32$^{G19V+T23N}$) *via* electroporation in *Ciona* fertilized eggs, as used for pTyr>Rab38^{G19V}. Pertubated larvae showed the three phenotypes that I observed using the single mutant: 56% of pertubated larvae had one well pigmented organs, while I did not observe the phenotype with one pigmented organs and the other not well melanized; 20% of larvae had not pigmented organs in the sensory vesicle while 24% of larvae seemed to have two normal pigmented sensory organs (Fig. 3.31).
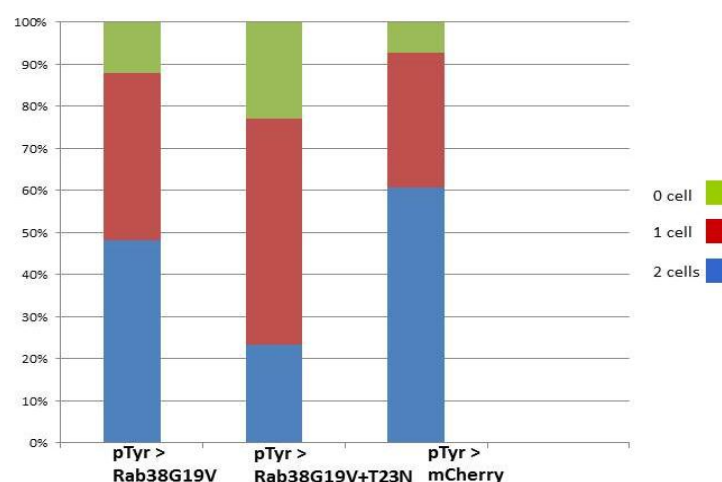
Fig.3.31 Percentage mutated Ci-Rab38/32 transgenic larvae. pTyr>Rab38^{G19V}, pTyr>Rab38$^{G19V+T23N}$ transgenic larvae and the control larvae, electroporated with pTyr>mCherry, showing no pigment cells (green), one pigment cell, either otolith or ocellus (red), or two pigment cells (blue). Within each column is the combined number of larvae counted during at least three trials; n>150 embryos scored for transgene expression.

Although the percentage of mutants is higher than the single mutant, there is still a small percentage of larvae appearing with no alteration in the two sensory organs. Nevertheless it's important to consider that the enhancer element present in the *cis*-regulatory region of *Ci-tyrosinase* is able to drive expression of the reporter gene mCherry in just 50% of electroporated embryos (Esposito et al., 2012); thus in order to evaluate how many larvae with two normal sensory organs are electroporated and to follow the formation of pigment cell lineage in mutant animals, I decided to electroporate the double mutant in switch I region together with the construct containing *pTyr* > mCherry.

As excepted, I observed a high reduction of pertubated larvae showing two pigmented sensory organs is drastically reduced (5%) (Fig. 3.32 A,D), demonstrating that the 24% of larvae electroporated just with the double mutant were wild type animals, not electroporated; moreover I observed an increase of larvae with no pigmented sensory organs (52%) (Fig. 3.32 A,C) while 46% of pertubated larvae showed just one single pigmented organs (Fig. 3.32 A,B).
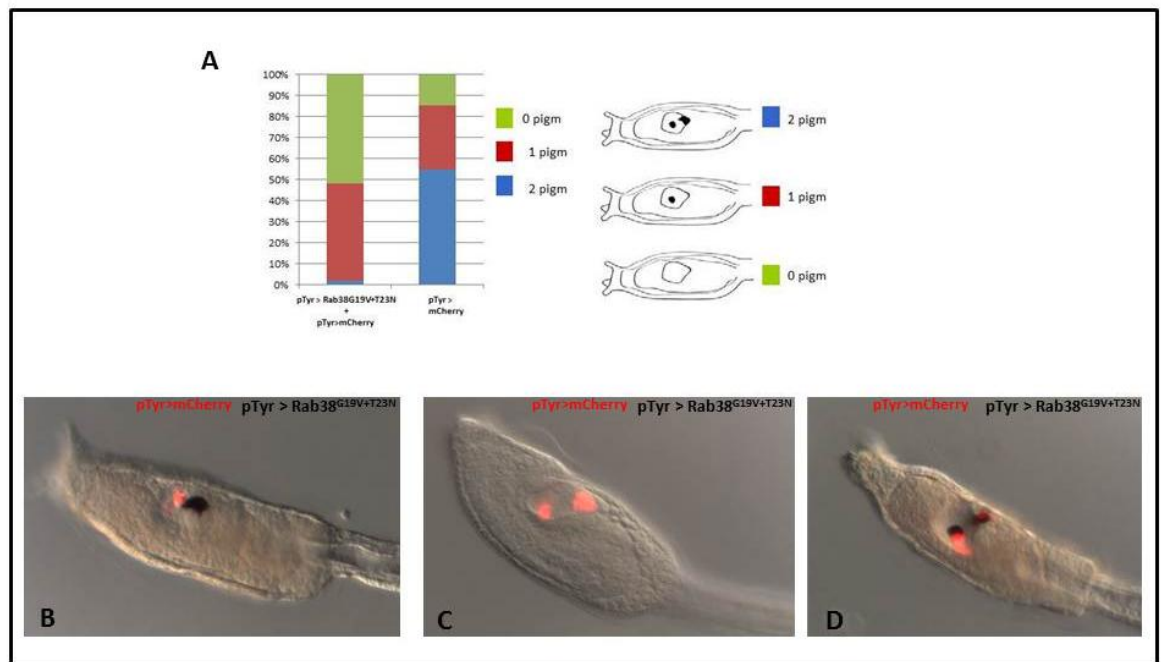
Fig.3.32 Ci-Rab38/32 is required for pigment cell formation (A,D). (A). pTyr>Rab38$^{G19V+T23N}$ transgenic larvae and the control larvae, electroporated with pTyr>mCherry, showing no pigment cells (green), one pigment cell, either otolith or ocellus (red), or two pigment cells (blue), as illustrated schematically on the right. Within each column is the combined number of larvae counted during at least three trials; n>150 embryos scored for transgene expression. (B-D) transgenic larvae co-electroporated larvae with pTyr>Rab38$^{G19V+T23N}$ + pTyr>mCherry, show defects in melanisation of pigment cell in the sensory vesicle, with 46% of pertubated larvae showing one single pigmented organs and 52% lacking of pigment cells in the sensory vesiscle. Lateral view, anterior towards the left.

3.10.3 CiRab38/32/rab-rp1/ltd promoter analysis

The non coding region that separate the ATG start codon in the first exon from the next

upstream gene is long 729 bp; in order to check if this region was able to recapitulate *Ci-Rab38/32/rab-rp1/ltd* gene expression in *C. intestinalis* embryos, I took advantage from GFP reporter gene expression assays. By PCR, I amplified a ~700 bp genomic fragment (named pRab38A) extending from the position −729 to −31 from the translation start site of the gene *Ci-Rab38/32/rab-rp1/ltd* (Fig. 3.33); this fragment was cloned into the vector pSP72 upstream the beta Globin minimal promoter to drive GFP expression. The resultant vectors, pSP72 pRab38A>GFP has been used to perform transgenesis via electroporation in *C. intestinalis* embryos.

The embryos have been observed from the neurula until the larval developmental stages, in fluorescent light microscopy. 50% of transgenic embryos showed GFP reporter expression in pigment precursor cell lineage, starting from the late tailbud stage to larval stage (Fig. 3.35 A,C,D). No evident GFP expression has been detected at the neurula stage, that correspond to the developmental stage where the endogenous transcript of *Ci-Rab38/32/rab-rp1/ltd* starts to be produced in pigment cell precursor blastomeres (a9.49s). Nevertheless, by WMISH assay, it has been possible to reveal the presence of reporter gene transcripts (GFP) in a9.49 blastomeres at the neurula stage. At the tailbud stage, GFP signals appears in the CNS in two of the four cells that correspond to PCPs, whereas the *Ci-Rab38/32/rab-rp1/ltd* transcript is present in all the four pigment cell lineage blastomers. At larval stage, GFP expressions have been very clearly detected around the pigmented cells of the two sensory organs, the otolith and the ocellus.

The GFP expression in the transgenic tailbud embryos reproduces partially the endogenous expression pattern of the *Ci-Rab38/32/rab-rp1/ltd* transcript, because no signals was detected at neurula stage; the slight delay in the visualization of the reporter gene in the a9.49 cell pair is likely due to a delay in the accumulation of the fluorescent protein in these cells, and not to loss of driving elements in the regulatory region taken into account. Furthermore the GFP expression has been detected just in two out the four cells that correspond to PCPs at tailbud stage. No expression has been detected for *Ci-Rab38/32/rab-rp1/ltd* transcript at larval stage, while GFP signals are present in both sensory organs in the larval sensory vesicle and continue to be stable also at late larval stage, when the animal starts to go in metamorphosis.
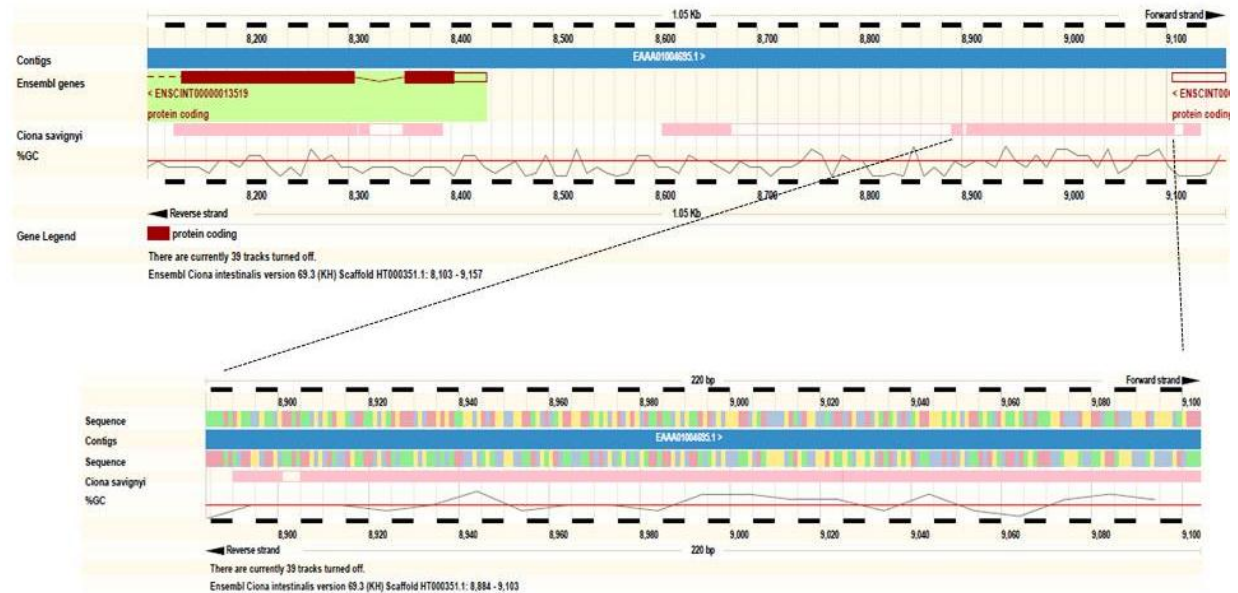
Fig. 3.33 *Ci-Rab38/32/rab-rp1/ltd cis*-**regulatory region and the fisrt conserved region.** *Ci-Rab38/32/rab-rp1/ltd* promoter located in the *Ciona*/Scaffold_ HT000351.1:7,483-8,434 and extending from the position −729 to −31 from the translation start site. Conserved transcript region with *C.savignyi* are evidenced as pink boxes; The Ciona savignyi reference sequence is on scaffold reftig_112:54031-54264. The first conserved region is located in the region between 9103 - 8884 of the scaffold (220bp) extending from position -690 to -450 from the ATG; the colored boxes represent the four nucleotide A (green), C (blue), G (yellow) and T(pink); the black picks represent the %GC in the transcript. Adapted from Ensembl release 70 - January 2013 version of *C. intestinalis* genome assembly v 2.0 .

In the promoter region of Ci-Rab38/32/rab-rp1/ltd, two conserved regions are present: the first is long 220bp, it is located in the region between 9103 - 8884 of the scaffold and corresponds to a region extending from position -690 to -450 from the ATG (Fig. 3.33).

The second conserved region is long 67bp, it is located in the region between 8674 -8606 of the scaffold and corresponds to a region extending from position -297 to -270 from the ATG (Fig. 3.34).
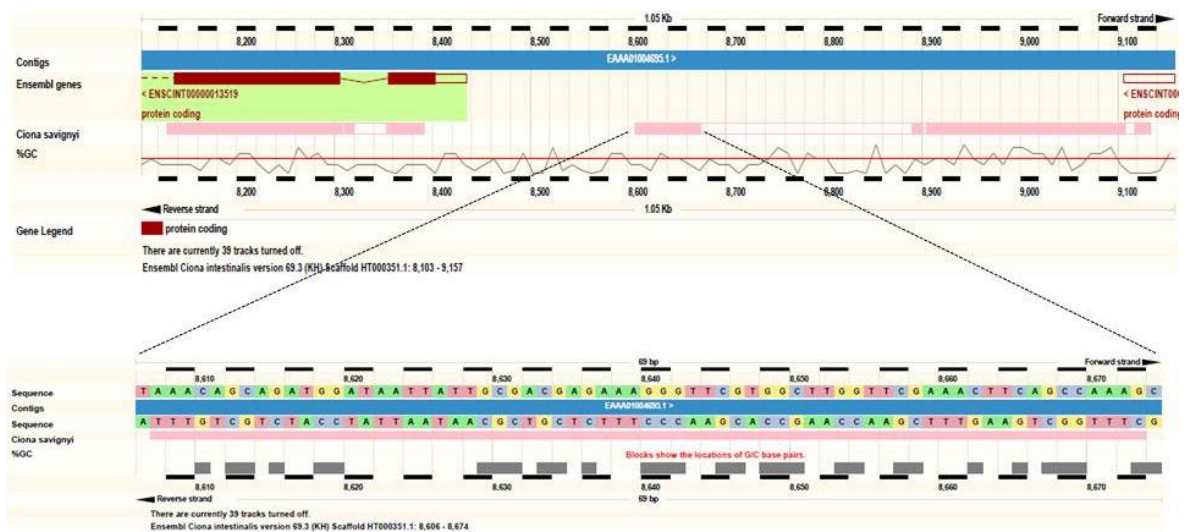
Fig. 3.34 *Ci-Rab38/32/rab-rp1/ltd cis*-**regulatory region and the second conserved region.** The second conserved region is located in the region between 8674 -8606 of the scaffold (67bp) extending from position -297 to -270 from the ATG; the colored boxes represent the four nucleotide A (green), C (blue), G (yellow)

and T(pink); conserved transcript region with *C.savignyi* is evidenced as pink boxes; the black picks represent the %GC in the transcript. Adapted from Ensembl release 70 - January 2013 version of *C. intestinalis* genome assembly v 2.0 .

To better narrow the region containing the putative enhancer elements responsible for *Ci-Rab38/32/rab-rp1/ltd* tissue specific expression, I decided to continue the promoter analysis subdividing the construct of 729 bp; taking into account the results of the C. intestinalis and *C. savignyi* genomic comparison, I subdivided the pRab38A in two fragments: one construct, pRab38B, contains the region extending from position -716 to -515 from the ATG and include the first conserved region, while the second construct, pRab38C, contains the second conserved region and it includes a region that extend from position -413 to -54 from the ATG (Fig. 3.35A-B).

The two constructs amplified by PCR with specific oligonucleotides has been cloned into the vector pSP72 upstream the beta Globin minimal promoter to drive GFP expression and assayed by electroporation experiments to verify the presence of positive *Ci-Rab38/32/rab-rp1/ltd* regulatory elements. Embryos at tailbud and larva stages have been observed with fluorescent light microscope to visualize the expression of GFP reporter gene. As for pRab38A , no GFP expression has been detected at the neurula stage for both the two constructs. For pRab38B I never observed reporter gene activation in electroporated embryos, both at tailbud and larval stage, while 40% of embryos electroporated with pRab38C showed a clear fluorescent signal both at the tailbud and larva stage as detected with pRab38A(Fig. 3.35 E-F).

These data revealed that, despite a high degree of sequence conservation, no positive regulatory elements seem to be present in the first conserved region, while they could be present in the second conserved region that was able to drive GFP reporter expression in same territories of pRab38A. In order to find minimal enhancer element in the cis-regulatory region of pRab38C, I continued to subdivide this construct in two fragments: pRab38D, from -244 to -54 to ATG in the first exon and pRab38E, extending from -413 to -245 to ATG (Fig. 3.35A).

As performed for the other constructs, I amplified them and cloned into the vector pSP72 upstream the beta Globin minimal promoter to drive GFP expression and assayed by electroporation experiments to verify the presence of positive regulatory elements. Using pRab38D construct for electroporation, no specific fluorescent signals were detected both at tailbud and larva stage, while embryos electroporated with pRab38E construct gave a specific reported activity as obtained for prab38C at tailbud in two PCPs and larva stage, in pigmented sensory organs, but with a less percentage, just 20% of embryos show a positive GFP expressing profile (Fig. 3.35A, G, H).
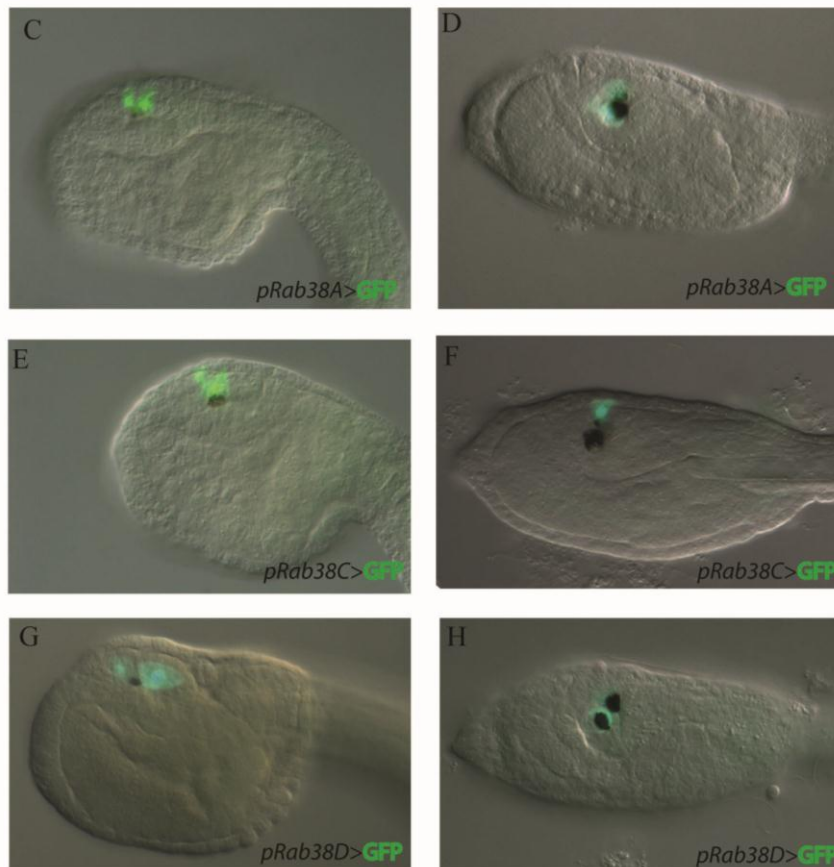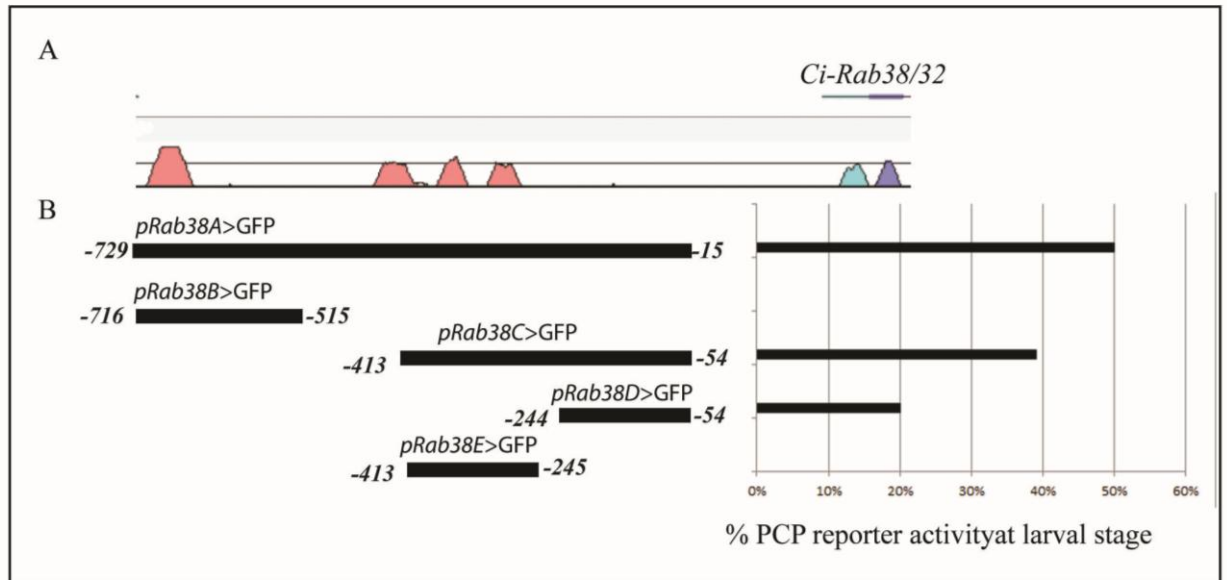
Fig. 3.35 *Cis*-regulatory region analysis of *Ci-Rab38/32/rab-rp1/ltd* (A) Vista alignment shows sequence conservation between *C. intestianlis* and *C.savignyi* in the 5' intergenic region; (B) Diagram of Ci-Rab38/32 reported constructs aligned with (A) and numbered by their distance from the translation start codon. Barplots indicate percentage of embryos showing PCP reported activity; (C-H) Transgenic embryos showing GFP expression in PCP at late tailbud stage (C,E,H) and pigment cell expression at larval stage (D,F,H). Lateral view, anterior to the left (C-H).

3.11 *In silico* analysis of the *Ci-Rab38/32/rab-rp1/ltd* promoter region

In order to identify putative binding sites for transcription factors that could be involved in the regulatory network of Ci-Rab38/32/rab-rp1/ltd, I next analysed in pRab38/32C (360bp) fragment using JASPAR database open source tool (http://jaspar.genereg.net). I selected the "JASPAR CORE Vertebrata" database that contains a curated, non-redundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for Vertebrates. The analysis revealed the presence of 356 Transcription Factor Binding Sites (TFBS) for in the sequence pRabB38/32C and 251 TFBSs in the sequence pRabB38/32D using a relative conservation profile score threshold of 80%.

Among all of the TFBSs in pRabB38/32C, I noticed the presence of six Ets binding sites (ETS1, matrix ID:MA0098.1) present on both DNA strands with a relative conservation profile score up to 92% (Table 2). Considering that FGF signalling is transmitted *via* the MAPK pathway, which activates the *Ci*-Ets1/2 transcription factor, it could be possible that *Ci-Rab38/32/rab-rp1/ltd* expression is in turn regulated by Ci-Ets1/2 resulting involved in the regulatory network controlled by FGF signalling in the *Ciona* pigmented organs formation.

Table 2 : predicted ETS1 binding sites in pRab38/32C							
Model ID	Model name	Score	Relative score	Start	End	Strand	Predicted site sequence
MA0098.1	ETS1	4.644	0.85	26	31	1	GTTCCA
MA0098.1	ETS1	3.822	0.81	44	49	-1	TTCCCG
MA0098.1	ETS1	5.777	0.90	45	50	-1	CTTCCC
MA0098.1	ETS1	5.896	0.92	179	184	1	TATCCA
MA0098.1	ETS1	4.241	0.83	223	228	1	AATCCA
MA0098.1	ETS1	3.658	0.80	329	334	-1	CTCCCT

Among the TFs identified, there were three TFBSs for ELK1, that belong to the Ets domain transcription factor family (Table 3). ELK-1 serves as an integration point for different mitogen-activated protein (MAP) kinase pathways. Phosphorylation of Elk-1 by MAP kinases triggers its activation (Yang et al., 2002). Ets family transcription factors are known effectors of FGF-MAPK signaling and can act as transcriptional repressors in various contexts (Hollenhorst et al., 2011; Oikawa and Yamada, 2003). Recently it is assessed that Ci-Elk-1, once activated by MAPK cascade, acts repressing *FoxC* expression in rows III/IV of the *Ciona* neural plate, a gene that is required together with *ZicL* for the anterio CNS formation (Wagner and Levine, 2012). It could be possible that ELK-1 could

be implicated in the regulation of *Ci-Rab38/32/rab-rp1/ltd* expressing acting as a repressor transcription factor.

Table 3: predicted ELK1 binding sites in pRab38/32C							
Model ID	Model name	Score	Relative score	Start	End	Strand	predicted site sequence
MA0028.1	ELK1	8.329	0.88	35	44	-1	GACCCTGAAG
MA0028.1	ELK1	6.753	0.83	41	50	1	GGTCGGGAAG
MA0028.1	ELK1	6.376	0.82	250	259	1	TAACCTGACG

Another interesting TFBSs present in the list of JASPAR analysis performed on pRabB38/32D fragment, are for MYC::MAX (matrix id: MA0059.1) and MAX (matrix id:MA0058.1) (Table 8); there are 9 binding sites for these matrices on both DNA strands with a relative conservation profile score up to 98% (Table 4). MAX transcription factor is annotated in *C. intestinalis* genome as *Ci-Mad/mnt/Rab24* that I selected from microarray data analysis as a new candidate FGF/MAPK/Ets target gene expressed in PCPs at tailbud stage (Fig. 3.23 F-F'').

MYC transcription factor is annotated as *Ci-myc* and on ANISEED database is reported that it is widely in the nervous system, both at neurula and tailbud stage (Imai et al., 2004). The Myc, Mad, Max network comprises a group of widely expressed transcription factors, which function in cell proliferation and differentiation. This network includes members of the Myc and Mad families, the Mad related protein Mnt, and Mga. All of these proteins

Table 4: predicted MAX and MYC::MAX binding sites in pRab38/32D							
Model ID	Model name	Score	Relative score	Start	End	Strand	predicted site sequence
MA0058.1	MAX	12.380	0.97	61	70	1	TACCACGTGG
MA0059.1	MYC::MAX	13.697	0.94	61	71	1	TACCACGTGGC
MA0059.1	MYC::MAX	13.250	0.93	62	72	-1	TGCCACGTGGT
MA0058.1	MAX	11.383	0.94	63	72	-1	TGCCACGTGG
MA0058.1	MAX	6.833	0.80	75	84	1	CAACGCGTGA
MA0058.1	MAX	7.063	0.81	136	145	1	GGTCACGAGG
MA0058.1	MAX	9.037	0.87	138	147	-1	AACCTCGTGA
MA0058.1	MAX	9.621	0.89	170	179	1	TCACACGTGA
MA0058.1	MAX	10.610	0.92	172	181	-1	AGTCACGTGT

possess basic helix–loop–helix zipper domains (bHLHZ), which mediate dimerization with the small bHLHZ protein Max (Grandori et al., 2000).

The Myc::Max complex ,as heterodimerize, binds E-box sequences and regulate transcriptionally cell growth, proliferation and differentiation (Amati et al., 1992). Ectopic expression of Myc TF in eye-antennal discs of Drosophila strongly interferes with their developmental program (Jiao et al., 2001), while overexpression of Mnt has strong effects on Drosophila wing and eye formation (Loo et al., 2005). Moreover considering that *Ci-myc* and *Ci-Mad/mnt/Rab24* show a decrease expression when comparing dnFGFR to control both at 8 and 12 hpf, it could be possible that these TFs cooperate in regulation of *Ci-Rab38/32/rab-rp1/ltd* activity in PCPs starting from neurula stage.

3.12 Identification of over represented transcription factor binding sites in the *cis*-regulatory regions of genes coordinately expressed.

When studying a cell signalling perturbation, a key challenge is to infer properties of a complex and dynamic Gene Regulatory Network (GNR) that control that cascade through the interactions between transcription factors (TFs) and their target genes. Two important components of the GNR are the DNA-binding trans-acting transcription factors (TF) and their corresponding transcription factor binding sites (TFBS) in the DNA sequence. Sets of proximal TFBSs that are sufficient to cooperatively mediate TF-regulated patterns of expression constitute *cis*-regulatory modules (CRM) (Kwon et al., 2012). Therefore, the activation or de-activation of TFs in specific signalling pathways triggers up- or down-regulation of their direct targets. In this context, transcriptomic approach, as DNA microarrays, generates list of different pattern of genes that share common properties as coordinated expression with common regulatory motifs in the upstream regulatory regions.

To advance the understanding of biological pathways and processes of a GNR, a key strategy is to identify over represented sequence motifs that are present in the *cis*-regulatory regions of genes coordinately expressed and select the mediating TFs for the observed co-expression patterns. Considering the assumption that some subset of the co-expressed genes is co-regulated by one or more common TFs, I employed a research using an open source tool, oPOSSUM, to identify statistically the over-representation of conserved TFBSs in the promoter regions of a set of co-expressed genes selected from my microarray analysis. oPOSSUM integrates a database of predicted and conserved TFBSs derived from JASPAR database (the version released in 2010) (Sandelin et al., 2004)and TFBS detection algorithms, using two statistical methods to determine the TFBS motifs that are over-represented in a Foreground set vs. the Background set sequences that are:

- Z-scores based on normal approximation to the binomial distribution that measures the change in the relative number of TFBS motifs in a foreground gene set compared with a background set

- Fisher scores based on a one-tailed Fisher exact probability assessing the number of genes with the TFBS motifs in the foreground set vs. the background set (Know et al., 2012). Using oPOSSUM system, two kind of analysis can be implemented: (1) a gene-based analysis where the count of the over-represented combination of motifs can be done for human, mouse, worm and yeast with a species-specific databases using a comparative genomics approach known as phylogenetic footpriting; (2) a sequence based analysis, where a target file containing sequences of the promoters is screened for the over-represented TFBS motifs with no conservation filtering applied or restriction in promoter size; in this case, it's suggested to introduce an own background sets with a prior biological evidence linking sequence composition to gene expression of a particular organism.

3.12.1 Selection of foreground and background datasets

The input file to implement oPOSSUM analysis is a set of co-expressed genes identified by an IDs and their correspondent putative regulatory regions. The foreground datasets derive from differential expression analysis on microarray data while the background dataset from the control samples. I applied for oPOSSUM analysis to identify new candidate TFs present in the GNR that control pigmentation in *Ciona* sensory organs and I decided to focus on gene list derived from comparison between dnFGFR to control samples at 12hpf; in particular I chose the 958 transcripts that have a significant statistical differential expression, defined by p value <0.05 and fold change > 1.5. In order to group genes present in the list for their expression values, I applied a *k* means clustering that is a method which aims to partition *n* observations into a fixed number (*k*) of clusters in which each observation belongs to the cluster with the nearest mean, such that the clusters are internally similar but externally dissimilar; the optimal choice of *k* should have a balance between maximum compression of the data using a single cluster, and maximum accuracy by assigning each data point to its own cluster. I tried with different values associating to *k*, from 1 to 8; I finally chose *k* = 4 to have the data set partitioned into four groups that show the optimal clustering pattern; the groups contains genes grouped upon their expression pattern, in particular Cluster 0 and Cluster 1 include most of the transcripts Up regulated, while Cluster 2 and Cluster 3 are composed by transcripts that are Down regulated (Fig. 3.36).
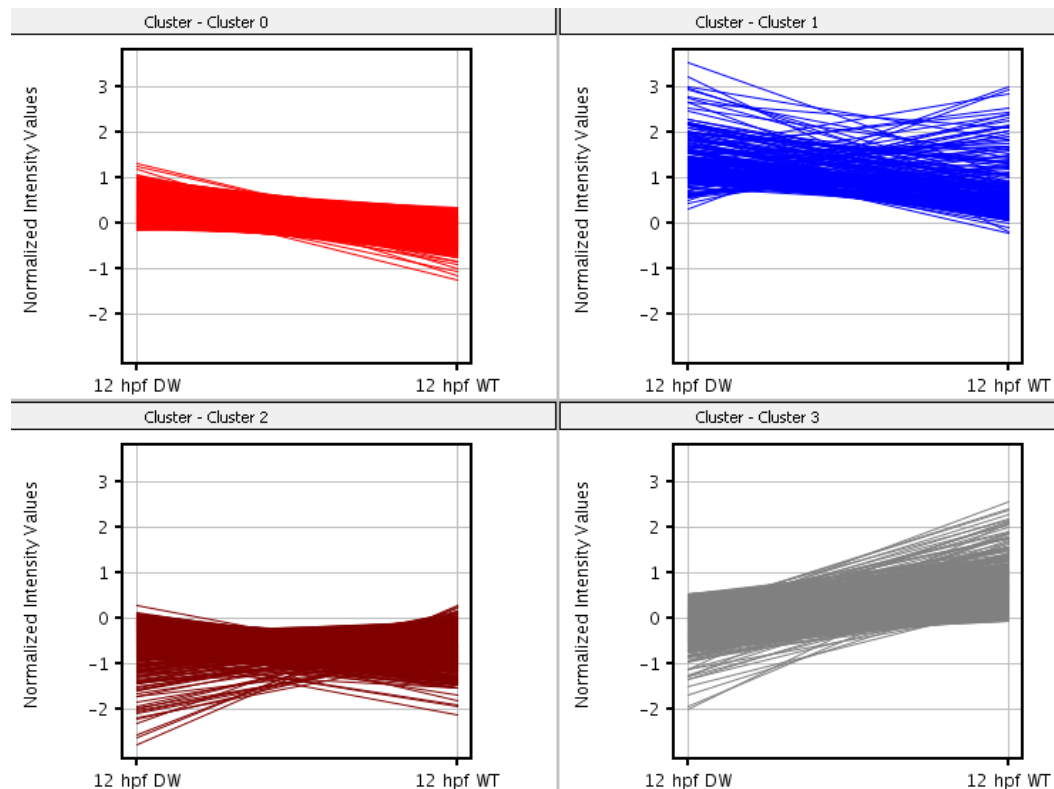
Fig. 3.36 Cluster set from *k* means clustering algorithm. the four clusters obtained from *k* means clustering are shown, the scheme with red lines are referred toCluster 0, blue to Cluster 1, brown to Cluster 2 and grey to Cluster 3; on both the *y*-axes, the normalized intensity averaged values of a transcript for the two conditions compared (dnFGFR and Control at 12 hpf , in the scheme referred as DW and WT) are reported. A colored line is plotted to link the values correspondent to the same transcript in the two conditions.

Once grouped the data set in clusters with co-expressed genes up or down regulated, I assessed oPOSSUM's performance. The system takes as input the list of identifiers of genes to be analysed, i.e. Ensembl IDs. Before to start with the analysis, I created an input file for foreground and background dataset where all genes have been annotated with Ensembl IDs and the correspondent sequences of putative regulatory elements upstream the TSS have been extracted. For the foreground dataset, the four clusters obtained from *k* mean clustering have been used, while the Background datasets were extracted from the control data of the microarray experiments; the number of genes selected for background size (~600 transcripts) was 2-fold greater than the foreground set. Transcript annotation and genomic sequences were retrieved from Ensembl v68, in particular I used Biomart (Kasprzyk, 2011), an open source tool available on Ensemble web site (http://www.ensembl.org/biomart/martview). To create the input file, the first step has been to select the probe set IDs of the genes from *Ciona* annotation file that I used as identifiers for the microarray and used Biomart system to retrieve cross-referenced gene annotation from the Ensembl database. I queried the database "Ensembl gene 70" and the dataset for *Ciona* "Ciona intestinalis gene (KH)"; to restrict the query, I filtered for "Affy CINT06a520380F probeste (ID)s" database that is the Affymetrix library file were each

probeset is associated to the Ciona genes. Moreover from Attributes, I chose to include in the output result sequences long 500bp from 5' flank coding region of each transcript. I decide to examine this size of the regulatory region because *Ciona intestinalis* genome is very compacted (~155 Mb in size) and most of minimal promoter regions, with putative regulatory elements isolated in *Ciona* genome, encompass 500 bp upstream the Transcription Starting Site, TSS. The BioMart software counts how many probeset IDs inserted are annotated from *Ciona* genome as an Ensembl IDs and gives as final result a FASTA formatted file with the transcript annotations from the Ensembl database; the corresponding intergenic sequences can then be scanned for TFBS hits in the oPOSSUM system (Fig.). Using Biomart tool I retrieved almost 50% of genes selected from *k* mean clustering in the four groups; in particular, for Cluster 0 270/476 Ensembl IDs with the 5' flanking region sequences were retrieved, for Cluster 1 569/886, for Cluster 2 574/1103 and for Cluster 3 202/424.

3.12.2 Application of oPOSSUM systems to microarray data

I assessed oPOSSUM's performance on the four sets of genes derived from *k* mean clustering. In order to apply a method for enrichment analyses for individual TFs, I firstly focused on Single Site Analysis (SSA) using a sequence-based analysis. Upon querying the system by a list of co-expressed genes with the upstream promoter sequences, oPOSSUM searches the matches to TFBS profile contained in JASPAR database and compute two statistics (Z-scores and Fisher scores) to measure the over-representation of TFBSs, returning a ranked list (Fig. 3.37). There are a number of parameters to specify in oPOSSUM sequenced-based analysis, including the Selection of transcription factor binding site from Jaspar matrices, matrix match threshold, promoter length and display option. All analyses were performed with the default parameters, in particular for selection of TFBS profiles with information content of at least 8 bits, 85% matrix match threshold and sequences with 500bp upstream the TSS. For result interpretation, I followed the suggestion by the authors of oPOSSUM that motifs with a Z-score exceeding 10 and a Fisher-score below 0.01 could be considered significantly over-represented (Ho Sui et al., 2007). However, I treated the target matrices as weakly over-represented when they satisfied only one of the above cutoffs.
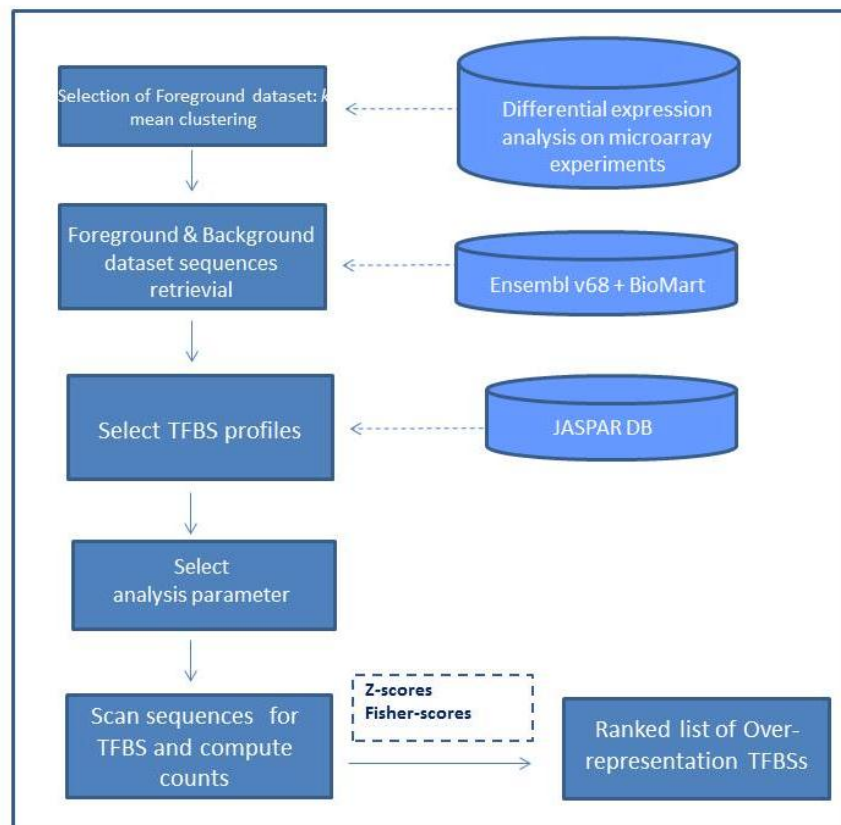
Fig. 3.37 Pipeline for the creation of inputs for oPOSSUM system to identify over-represented TFBSs.
See the text for a step-by-step description of the pipeline.

3.12.3 oPOSSUM results for Cluster 2 and Cluster 3

For Cluster 2, I assessed oPOSSUM's performance on 574 genes that showed decreased expression levels when dnFGFR is expressed at 12hpf (Table 5); I recovered ELF5 TFBSs as one of the most over-represented matrix ranked in the promoters of Cluster 2 genes by either Z scores and Fisher score; ELF5 (Ets Related Transcription factor 5) belongs to the Ets family of TFs, class Winged Helix-Turn-Helix, that recognizes the consensus DNA motif "(A/T)TCC". ELF5 binding sites were identified in 320 of the co-expressed genes in Cluster 2.

Table 5. Statically over-represented TFBSs in Cluster 2

TF	Family	Target seq hits	Background seq hits	Z-score	Fisher score
NFE2L2	Leucine Zipper	49	17	16.866	0.799
MEF2A	MADS	148	67	15.259	0.058
Spz1	Other	49	18	15.203	0.629
Lhx3	Homeo	149	63	15.1	0.165
ELF5	Ets	320	144	13.965	0.002
HNF1A	Homeo	60	25	13.313	0.317
NR2F1	Hormone-nuclear Receptor	24	9	12.607	0.529
FOXI1	Forkhead	407	176	11.646	0
Pax5	Homeo	7	2	9.571	0.622
FEV	Ets	255	127	8.342	0

TFBSs detected by oPOSSUM with the mostly highly ranked Z-scores

Considering that FGF signalling via Ets is the induction process involved in pigmentation of *Ciona* sensory organs, it's consistent that an Ets TF family member is a top-scoring matrix ranked for both F and Z scores. Besides ELF5, I also found another matrix belonging to Ets family TFs, FEV, that has binding motifs present on 255 sequences scanned; however this TF is ranked with a higher Fisher score, more than ELF5, but a lower Z-score (8.34), lower the threshold for significant over-representation. Cluster 2 contains some of the PCPs markers that I have identified from microarray experiments, as *Ci-BMP5/7*, *Ci-Lrp4*, *Ci-Mnt/Mad*, *Ci-slc45a2* (Fig. 3.23 B-G'') together with known melanogenic genes, such as *Ci-Tyrosinase* and *Ci-Tyrp1/2a*. As we have already demonstrated that melanogenic process is directly regulated FGF/MAPK/Ets signalling pathway, I could consider the over-representation of these two Ets TF family members as a positive control for the oPOSSUM performance.

In the list of the over represented TFBSS for Cluster 2, as determined by Z and F-scores cut off, I found a matrix for FOXI1, a member of the forkhead-box TF family that is linked to G-protein coupled receptor protein signaling pathway. FOXHI1 has a large number motifs in more than 400 sequences. In *Ciona* genome, the homolog of FOXI1 is annotated as *Ci-FoxIb*, a gene expressed in the nervous system at tailbud stage (Mazet et al., 2005). Moreover *Ci-FoxIb* shows an increased expression in microarray data with a fold change higher than 2 when comparing the control at 12hpf with control samples at 8 hpf. This variation is consistent with *Ci-FoxIb* expression pattern that starts to be detected at tailbud stage (12 hpf). These observations suggest that *Ci-FoxIb* is a good candidate involved in FGF signalling cascade that lead to *Ciona* pigmented sensory organs formation. Furthermore *Ci-FoxF*, another forkhead/winged helix transcription factor, has been described as a direct target of FGF signalling and is predominantly involved in the

regulation of heart cell migration (Beh et al., 2007); in particular, *Ci-FoxF* is up-regulated by FGF signalling, transduced via the MAPK pathway and Ets activation, and interfering with *Ci-FoxF* function leads to inhibition of heart precursor cells migration (Christiaen et al., 2008). Another matrix ranked in the top five enriched TFBSs is for LHX3, a Homeo domain TF. LHX3 is ranked in the top five just for the Z-score but not for F-score, therefore I considered this TF as weakly over-represented according to the statistical tests. oPOSSUM system gives the possibility to display the list of identifiers (Ensembl IDs) of the genes where TFBS for LHX3 are located. The next step has been to check "one-by-one" the Ensembl IDs to retrieve the *Ciona* gene names and the putative gene expression patterns whenever they are reported querying the ANISSED database. Interestingly the 90% of the 150 sequences are annotated in *Ciona* genome and among them I found melanogenic genes, such as *Ci-Tyrosinase*, *Ci-Tyrp1/2a*, *Ci-Tyrp1/2b*, and *Ci-Frizzled receptor*. This last factor is a receptor involved in Wnt pathway already demonstrated to cooperates in *Ciona* pigment cell induction (Squarzoni et al., 2011); among these 150 sequences, most of them have no expression pattern reported on ANISEED database except for *Ci-Ddx3x*, that is an ATP dependent RNA helicase widely expressed in *Ciona* nervous system at tailbud stage. Because nothing is known about these genes it could be very interesting to study their expression pattern or biological function in order to find new candidate genes involved in pigment cell formation.

Subsequently I submitted to a sequenced-based SSA the Cluster 3 list that contains most of genes showing a decreased expression values comparing dnFGFR to control at 12hpf; also in this case FEV, that belong to Ets TF family, is one of the top ranked matrix under the Z- and F-scores, which satisfies the thresholds for significant over-representation (Table 6).

Table 6. Statically over-represented TFBSs in Cluster 3

TF	Family	Target seq hits	Background seq hits	Z-score	Fisher score
SRF	MADS	4	1	15.657	1.79
Sox2	High Mobility Group	12	7	15.636	1.967
FEV	Ets	96	127	11.041	0.007
Spz1	Other	18	18	9.17	0.713
HNF1A	Homeo	23	25	9.101	0.528
EBF1	Helix-Loop-Helix	15	17	8.927	0.443
SOX9	High Mobility Group	112	152	8.747	0
HLF	Leucine Zipper	73	107	7.877	0.002
NR4A2	Hormone-nuclear Receptor	54	84	7.648	0.003
FOXO3	Forkhead	145	186	7.18	0

TFBSs detected by oPOSSUM with the mostly highly ranked Z-scores

In the ranked list for Z-score values for Cluster 3 analysis, I recovered SOX9 matrix, a TF that belongs to SRY type HMG box family. In human melanocytes SOX9 is expressed in neonatal and adult skin and regulates genes involved in pigmentation, such as *MITF*, *Tyrosinase* and *Tyrp*, leading to an increase in the expression of these genes that are key melanogenic proteins and finally to a stimulation of pigmentation (Passeron et al., 2007). In *Ciona* genome, SOX9 is annotated as *Ci-SoxC*, a TFs that starts to be expressed at tailbud stage in epidermis and nervous system, including PCP territories (Imai KS et al., 2004). SOX9 matrix has been identified in 110 sequences for Cluster 3 genes; I checked the corresponding *Ciona* names and I found some genes previously selected in my microarray data analysis for their biological function, such as *Ci-Fringe*, *Ci-Osbpl1a, Ci-Pla2g4b*, *Ci-Cadherin-related-7*, *Ci-Cav-a2d2*, *Ci-calmodulin*.

In microarray data, *Ci-SoxC* is not differentially expressed, in any statistical analysis performed, suggesting that this gene is not regulated by FGF cascade in PCPs but could in turn have a role in the regulation of other genes that are involved in the formation of pigmented sensory organs.

3.12.4 oPOSSUM results on Cluster 0 and Cluster 1

Cluster 0 and Cluster 1 lists present most of genes Up regulated when dnFGFR is compared to control samples at 12hpf.

The PPARG::RXRA matrix is found to be a top-scoring motif in the analysis performed on both Cluster 0 and Cluster 1; peroxisome proliferator activated receptors (PPARs) bind preferentially to DNA as heterodimers with a common partner, retinoid acid X receptor (RXR), to regulate transcription. In *Ciona* genome both receptors are present. PPARG::RXRA bound regions were present on 11 promoter from Cluster 0 analysis and 13 sequences from Cluster 1; interestingly among sequences of Cluster 1 is present *Ci-Bcmo*, which is involved in retinoic acid biosynthesis (Cui and Freedman, 2009), while among sequences of Cluster 0 are present genes, as *Ci-cher*, *Ci-Calmodulin*, *Ci- dpld* and *Ci-mmd*, expressed in the nervous system. Using the 569 genes of Cluster 1, I recovered RUNX1 as the top ranking for Z-scores (Table 7).

Table 7. Statically over-represented TFBSs in Cluster 1

TF	Family	Target seq hits	Background seq hits	Z-score	Fisher score
PPARG::RXRA	Hormone-nuclear Receptor	11	2	23.61	1.342
Pou5f1	Homeo	43	14	18.099	0.998
RUNX1	Runt	225	100	14.691	0.034
Zfp423	BetaBetaAlpha-zinc finger	19	7	12.772	0.551
Nkx2-5	Homeo	549	212	10.702	0.008
HOXA5	Homeo	544	210	9.674	0.035
MYC::MAX	Helix-Loop-Helix	28	11	9.275	0.472
YY1	BetaBetaAlpha-zinc finger	435	189	9.255	0
NKX3-1	Homeo	443	196	9.131	0
MZF1_5-13	BetaBetaAlpha-zinc finger	124	72	8.666	0

TFBSs detected by oPOSSUM with the mostly highly ranked Z-scores

In *Ciona* genome, RUNX1 is annotated as *Ci-Runx*, a TF belonging to Runt TF family; RUNX1 TFBSs are present in 225 sequences. Interestingly I found that TFBS for RUNX1 are present on promoter region of a gene I previously selected for its anterior CNS expression pattern in *Ciona* tailbud, *Ci-Pax3/7*; moreover, among the 225 sequences, I found *Ci-Bcmo* that is a gene encoding for β-carotene 15,15′-monooxygenase, a key enzyme involved in the metabolism of plant beta-carotene to retinal (Park et al., 2008). During *Ciona* embryogenesis, *Ci-Bcmo* starts to be expressed in the nervous system of neurula stage and in the anterior and posterior sensory vesiscle territories, including PCPs, from tailbud to larval stage (Nakashima et al., 2003).

Another prominently enriched profile in oPOSSUM result is represented by NK3-1, that has a large number of TF bound region (443 sequences) in the promoters of Up regulated genes. In *Ciona* genome NK3-1 is annotated as *Ci-Bapx1*, a Nirenberg and Kim homeobox gene member of the NK TF family, that in human has an established role in skeletal development and controls the expression of several molecular markers required for normal chondroblast differentiation, such as SOX9 and Fgfr3. In microarray data, *Ci-Bapx1* has a robust increased expression when dnFGFR is compared to control both at 8 and 12 hpf and it could be a good candidate TF that have a role in the regulation of genes in the GNR of FGF signalling cascade.

The analysis on the 270 genes of Cluster 0, revealed MZF1_5-13 as one of the most over-represented ranked for Z scores (Table 8); this TF belong to the Krüppel family of zinc finger proteins and in *Ciona* genome is annotated as *Ci-znf061*, a gene that starts to be expressed in central nervous system at tailbud stage (Imai et al., 2004). MZF1_5-13 TFBSs

are present on 77 sequences and among them I found genes that are widely expressed in *Ciona* nervous system, as *Ci-CRYBA1*, *Ci-ADAMTSL1*, *Ci-cher* and *Ci-neurocalcin*. Considering that MZF1_5-13 TFBS are also over-represented on promoters of genes previously selected from microarray analysis, such as *Ci-FoxHa*, *Ci-slc7a9*, *Ci-COE*, *Ci-EphrinA-c* that share an anterior expression in the CNS, it could be possible that MZF1_5-13 is involved in the simultaneously regulation of these co-expressed genes.

Table 8. Statically over-represented TFBSs in Cluster 0

TF	Family	Target seq hits	Background seq hits	Z-score	Fisher score
PPARG::RXRA	Hormone-nuclear Receptor	13	2	36.139	4.47
ESR2	Hormone-nuclear Receptor	6	1	25.494	2.22
Ar	Hormone-nuclear Receptor	4	1	17.41	1.32
MZF1_5-13	BetaBetaAlpha-zinc finger	77	72	13.459	0.098
NR3C1	Hormone-nuclear Receptor	14	8	11.709	1.21
NFE2L2	Leucine Zipper	25	17	11.072	1.002
NR4A2	Hormone-nuclear Receptor	95	84	9.908	0.166
Zfp423	BetaBetaAlpha-zinc finger	11	7	9.591	0.875
Pou5f1	Homeo	18	14	8.838	0.589

TFBSs detected by oPOSSUM with the mostly highly ranked Z-scores

oPOSSUM system has been a powerful tool to seek insight into GNR to identify new candidate TFs that may acting in a peculiar biological context. Considering the TFs identified as over-represented in genes differentially expressed from microarray experiments are recovered computationally, it could be very interesting to better analyse their involvement in the GNR that lead to pigmentation in *Ciona* sensory organs or in the formation of anterior CNS.

CHAPTER 4

CONCLUSIONS

4.1 FGF signalling in Central Nervous System differentiation

The FGF signaling plays an important role in neural induction and subsequent development throughout the vertebrates (Bertrand et al., 2003; Dono, 2003; Ribisi et al., 2000; Streit et al., 2000). In chick embryos, FGF signalling is implicated in the earliest inductive events of neural tissue, which occur before the onset of gastrulation (Wilson et al., 2000). Upon binding to its receptor, FGF activates the Ras/MEK/ERK signalling cascade through a receptor tyrosine kinase, FGFR (Szebenyi and Fallon, 1999). This early neural induction, which is implicated to generate tissue of anterior character, is followed by a stabilisation step to give committed neural precursors (Vasiliauskas and Stern, 2001). Finally, a posteriorisation step is required to generate posterior neural tissue, where pigmented cell precursors are located.

Likewise, the FGF signaling pathway also functions in neural development in ascidians. In *Ciona*, neural induction begins with *Otx* expression at the 32-cell stage in the anterior neurectoderm via GATA and Ets transcription factors in response to FGF signalling (Bertrand et al., 2003). The requirement for FGF signaling persists in later development, when it is required for patterning posterior components of the nervous system (Imai et al., 2009; Stolfi et al., 2011). Thus, FGF signaling acts at both early and late stages of *Ciona* nervous system development.

Recently it has been assessed that FGF/Wnt signaling cascade have a crucial function in *Ciona* pigmented sensory induction and formation (Squarzoni et al., 2011); these signaling pathways cross talk by a direct transcriptional regulation exerted by the FGF effector Ets, on the Wnt downstream gene *Ci-Tcf*. Nevertheless the FGF downstream targets, involved in pigment cell precursors formation, are not yet characterized in *Ciona*. In the frame of my PhD project, I focused the attention on extending the characterization of FGF cascade with the final aim to depict the gene regulatory network that controls CNS differentiation and pigmented cell specification in *C. intestinalis*. To this aim, I employed a transcriptomic approach, using a microarray analysis, by exploiting a targeted interference of the FGF signalling together with the enrichment of cells of interest by Fluorescence Activated Cell Sorted (FACS). The application of GeneChip microarray strategy associated to cell sorting has facilitated the identification of novel candidates specifically expressed in pigment cell lineage and contributed to the selection of unknown functional target involved in the gene regulatory networks controlling *Ciona* CNS differentiation.

The complex gene regulatory architecture underlying CNS formation has been difficult to characterize in vertebrate. Due to its simplicity *C. intestinalis* represent a potent model for systems level analysis of chordate nervous system differentiation providing critical insights into the related but more complex vertebrate nervous system regulatory networks.

Target molecular interference of FGF signalling represent a powerful system for deciphering the downstream target gene network that could be directly or indirectly regulated by the cascade. FGF signaling can be either potentiated or disrupted leading to a strongly impaired pigment cell specification. To drive a target interference in PCPs, I employed a molecular manipulation of FGF function, using a pigment cell precursors specific enhancer element, *pTyrp1/2a*, to selectively induce the expression of an inhibited condition (dnFGFR) and a hyper activated condition (EtsVp16). Furthermore to provide a temporal insight in the initial and terminal FGF signaling regulatory network, the pertubated conditions have been targeted at two developmental stages: 8hpf (neurula stage) corresponding to a FGF-mediated early induction, and 12hpf (tailbud stage) where pigment cells are already fate restricted.

4.2 Computational approach for a deep insight in *Ciona* CNS formation and differentiation

A transcriptomic approach using microarray analysis has been a valuable resource to produce lists of differentially expressed genes and to elucidate the FGF/MAPK/Ets regulatory architecture providing temporal specificity on the gene regulatory network controlled by FGF signaling pathway.

The enormous amount of information generated by high throughput profiling technologies offers previously unknown scientific opportunities and new technical challenges. The development of microarray technology allows the simultaneous measurement of the expression of many thousands of genes. The information gained offers an unprecedented opportunity to fully characterize biological processes, as the CNS formation and specification. However, to take full advantage of the huge opportunities offered by this information, it is necessary to manage, integrate and interpret these data correctly with appropriate bioinformatic knowledge. Easy, comprehensive approaches are needed to make sense of microarray data.

Microarray data analysis begins with normalization, which attempts to reduce the experimental variability across different array spots while maintaining biological variability. Several statistical methods have been developed to normalize microarray data and enable further statistical analyses to address biological variability. Statistical analysis

provide a list of gene differentially expressed when comparing conditions to control samples. The next question is what to do with such a list, other than look for genes that are considered interesting?

One method of analysing microarray data, that is becoming popular is pathway analysis, also known as functional enrichment. This integrates the normalized array data and their annotations, such as metabolic pathways and gene ontology functional classifications. There are many sources of pathway and functional information, which can be either generic or species specific.

The Gene Ontology project (GO) (http://www.geneontology.org) (Harris, M.A. et al. 2004) classifies genes into a hierarchy, placing gene products with similar functions together. The GO project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The project began as a collaboration between three model organism databases, FlyBase (Drosophila), the Saccharomyces Genome Database (SGD) and the Mouse Genome Database (MGD), in 1998. Since then, the GO Consortium has grown to include many databases, including several of the world's major repositories for plant, animal and microbial genomes.

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) (www.genome.jp/kegg) (Kanehisa et al., 2004) provides searchable pathways. GenMAPP (Dahlquist et al., 2002) (http://www.genmapp.org) provides an image of a pathway that is annotated with accession numbers. Software is available to view gene expression data either on a pathway (e.g. KEGG and MAPPFinder) or within the gene ontology (e.g. AmiGO) (http://www.godatabase.org).

However pathways and functional analysis needs different types of annotations, such as Affymetrix probe IDs, GenBank, RefSeq and SwissProt accession numbers.

One consistent limitation for my analysis has been that *C. intestinalis* GeneChips are custom Affymetrix microarrays; thus I needed to create a probe annotation file to associate every probeset to a specific gene and its functions, where the nucleotide sequences of probes are retrieved from Affymetrix library file (CINT06a520380F.sif). *C. intestinalis* genome is not a well-annotated genome and to accumulate as many information as possible, the transcript models and annotations were downloaded from many sources such as the ANISEED database (Tassy et al., 2010) on March 2011: 1) JGI version 1, 2) KYOTOGRAIL2005, 3) KH and 4) ENSEMBL. However I collected a partial annotation for *Ciona* genome, regarding biological functions, GO Ids and no specific Pathway are already available for this organism. Due to this limitation, I used the homologies of *Ciona* genes to well annotated genome of other organism, vertebrate and invertebrate, to enrich

all the information about *Ciona* genes not annotated, as gene functions, protein structures and domains and protein-protein interactions.

Gene expression profiles can be used to infer previously unknown transcriptional regulatory interaction among thousands of genes, via systems biology 'reverse engineering' approaches (De Cegli et al.,2013). The aim is to infer, or 'reverse-engineer', from gene expression data, the regulatory interactions among genes using computational algorithms. There are two broad classes of reverse-engineering algorithms (Faith and Gardner, 2005): those based on the 'physical interaction' approach that aim at identifying interactions among transcription factors and their target genes (gene-to-sequence interaction) and those based on the 'influence interaction' approach that try to relate the expression of a gene to the expression of the other genes in the cell (gene-to-gene interaction), rather than relating it to sequence motifs found in its promoter (gene-to-sequence) (Bansal et al., 2007). Regarding my microarray data analysis, I applied for both computational approaches; in collaboration with G. Gambardella from TIGEM, we 'reverse engineered' a pigment cell lineage-specific transcriptional network from my microarray data set to identify new potential candidates involved in FGF signaling regulatory architecture; on the transcriptional co-regulatory network created, a neighbors analysis have been employed to find genes with a significant value of Spearman Correlation Coefficient, thus co-expressed with melanogenic genes, such *as Ci-Tyrosinase*, *Ci-Tyrp1/2a*, *Ci-Tyrp1/2b*, *Ci-Tcf/LEF1*. Among the gene lists produced, I selected several good candidates, such as *Ci-slc45a2*, *Ci-Hunchbacklike*, *Ci-Mad/Mnt*, *Ci-Exosc7* and *Ci-Rab38/32*, for the biological validation that revealed their involvement in the gene regulatory network controlled by FGF signaling in PCPs. However further biological analyses are necessary to delineate the roles of these genes and the other candidates identified in FGF signaling regulatory network.

Clustering algorithm, such as hierarchical and *k* means clustering, are consolidated methods to group or classify samples and genes according to similar expression patterns. Once identified cluster of genes that have a similar expression values, there are several open source tools available to identify interactions among transcription factors and their target genes. Among these, I chose to use oPOSSUM software (Kwon et al., 2012) to select statistically the over-representation of conserved transcription factor binding sites in the promoter regions of a set of co-expressed genes. I assessed oPOSSUM's performance after selected from microarray data set groups of genes with a similar expression using *k* mean clustering analysis.

oPOSSUM system has been a powerful tool to seek insight into FGF signaling gene regulatory network to identify putative candidate transcription factors (TF) that are involved in PCPs specification or CNS differentiation. Considering that these TFs

identified have been recovered computationally, further analyses are necessary to prove their biological involvement in the FGF signaling cascade pathway.

Transcript profiling offers the largest coverage and a wide dynamic range of gene expression information and can often be performed genome wide. Microarrays are currently most popular for transcript profiling, they are established tools that the research community is familiar with, plus the bioinformatics pipelines for array data analysis are mature.

However several next-generation sequencing (NGS) platforms are recently harnessing the power of massively-parallel short-read DNA sequencing to digitally interrogate genomes on a revolutionary scale. For examples serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS) are well-established alternatives to microarrays. Both techniques can be superior to microarrays because they do not depend on prior probe selection. More recently, direct sequencing of transcripts by high-throughput sequencing technologies (RNA-Seq) has become an additional alternative to microarrays and is superseding SAGE and MPSS (Busch and Lohmann, 2007). Like SAGE and MPPS, RNA-Seq does not depend on genome annotation for prior probe selection and avoids biases introduced during hybridization of microarrays. On the other hand, RNA-Seq poses novel algorithmic and logistic challenges, and current wet-lab RNA-Seq strategies require lengthy library preparation procedures. Because of their robust sample processing and analysis pipelines, often microarrays are still a preferable choice for projects that involve large numbers of samples for profiling transcripts in model organisms with annotated genomes. Microarrays still provide a valuable resource to produce a huge amount of data from which extract lists of differentially expressed genes that could represented a starting point for a deep sequence analysis. There is still the possibility to imagine a powerful symbiosis between microarrays and NGS technologies. Arrays may be best suited in classifying cohorts of samples and once samples of interest are defined, NGS could be used to provide comprehensive deep sequence analysis of either genomic DNA to identify mutations, or RNA to report differences at the transcriptome level.

4.3 FGF signaling is required for setting the anterior-posterior CNS patterning in *Ciona*

In vertebrates, neural territories are composed of mixed fate cell populations, which segregate and form boundaries according to fate at comparatively later developmental stages in response to diverse signals, including FGFs (Cajal et al., 2012; Sanchez-Arrones et al., 2012; Toro and Varga, 2007).

Ciona embryos develop from defined cell lineages, and the gene expression patterns, that underlie important embryological boundaries, form relatively early in development. *Ciona* PCPs (a9.49 cell pairs) arises from the a8.25 blastomers that divide one time into the a9.49s and a9.50s (Fig. 4.1); cells in rows III-VI derive from the anterior ectodermal lineage known as the a-line. Progeny of row III contribute to the anterior sensory vesicle whereas cells of row IV contribute to the stomodeum as well as the anterior brain.

At late gastrula stage, a9.49s are in close proximity to two sources of FGF diffusible signaling molecules, Ci-FGF8/17/18 and Ci-FGF9/16/20 that are released from the lateral and medio-central nerve cord precursor cells of the underneath row, respectively (Hudson et al., 2007b). At the mid-neurula stage, the a9.49 cell pair division gives rise to four cells the a10.97s and a10.98s, that at later stage intercalate each other lining up dorsally in the forming neural tube; meantime cells derived from a9.50s division are located more anteriorly than the lined PCPs. Morphological embryonic rearrangement leads the a10.98s in a more rostral position, outdistancing them and a9.99 and a9.100 cell pairs from FGF sources, while the a10.97s remains within the FGF signaling range. Among these four PCP cells, only the a10.97s cells develop into *C. intestinalis* pigment cells while the a10.98s form part of the right-dorsal sensory vesicle epithelium. Despite their common origin, the different cell fate of PCPs is maybe due to the diverse FGF responsiveness and distance from FGF sources of a10.97 and a10.98 cell pairs. Indeed it is noteworthy that FGF molecules act at a very short range from the target cells (Tassy et al., 2006) and FGF induction is dependent on the distance between the source and the target cell. During the last a9.49 cell pair division, FGF responsiveness could be differentially inherited by the a10.97 and a10.98 cell pairs, contributing to their cell-specific fate diversification.

From microarray analysis I identified new marker genes for pigment cell lineage, specifically expressed in PCPs; it appears that some genes are expressing exclusively in the a10.98s, such as *Ci-BZRAP*, *Ci-Slc45a2*, while other are expressed just in the a10.97s, like *Ci-Doc2a*, supporting the idea that the four PCPs have a cell specific fate expressing different gene patterns (Fig. 4.1).

Furthermore from microarray data analysis it appears that FGF signaling is required for setting the anterior-posterior CNS patterning in *Ciona*. I mainly observed that when FGF is blocked in PCPs, the most anterior cells change their fate, expressing genes that are normally present in the anterior region of CNS, such as *Ci-Six3/6, Ci-Pax3/7, Ci-Rgr/Nut2, Ci-Neurogenin, Ci-FoxHa, Ci-EBF/COE, Ci-SLC7A9, Ci-Macho1*. At tailbud stage *Ci-Six3/6* is expressed in a wide region of the anterior CNS with the most posterior expression detected adjacent but not overlapping with PCPs, at the boundary line with a10.98s as revealed by double WMISH with *Ci-Tyrp1/2a* (Fig. 3.37). Additionally, when dnFGFR is

expressed at tailbud stage I observed an ectopic expression of *Ci-Six3/6* in a posterior region of the CNS, where the a10.97 cells are located (Fig. 3.38), supporting the hypothesis that FGF signalling is implicated in the repression of the anterior CNS specification in the pigment cell lineage.

In summary, the boundary between rows III and IV of the *Ciona* neural plate is important because it delineates the anterior sensory vesicle containing pigmented sensory organs from the anterior brain region and stomodeum. Microarray data analysis reveals that FGF-MAPK-Ets1/2 signaling is required to establish this boundary and that it does so by repressing a sub set of genes specifically expressed in the anterior CNS in the pigment cell lineage. Disruption of FGF signaling leads to an ectopically expression of these genes in the most anterior cells representing PCPs at tailbud stage.

In conclusion, my results indicate that FGF signaling is required for setting the anterior-posterior differentiation in *Ciona* CNS and it is necessary for delineate the posterior boundary of pigmented cell lineage expression domain; FGF cascade is also required to establish the proper identify of cells in row III and row IV, determining the fate of a9.49s as PCPs and contributing to cell-specific fate diversification of a10.98s and a10.97s in later stages.
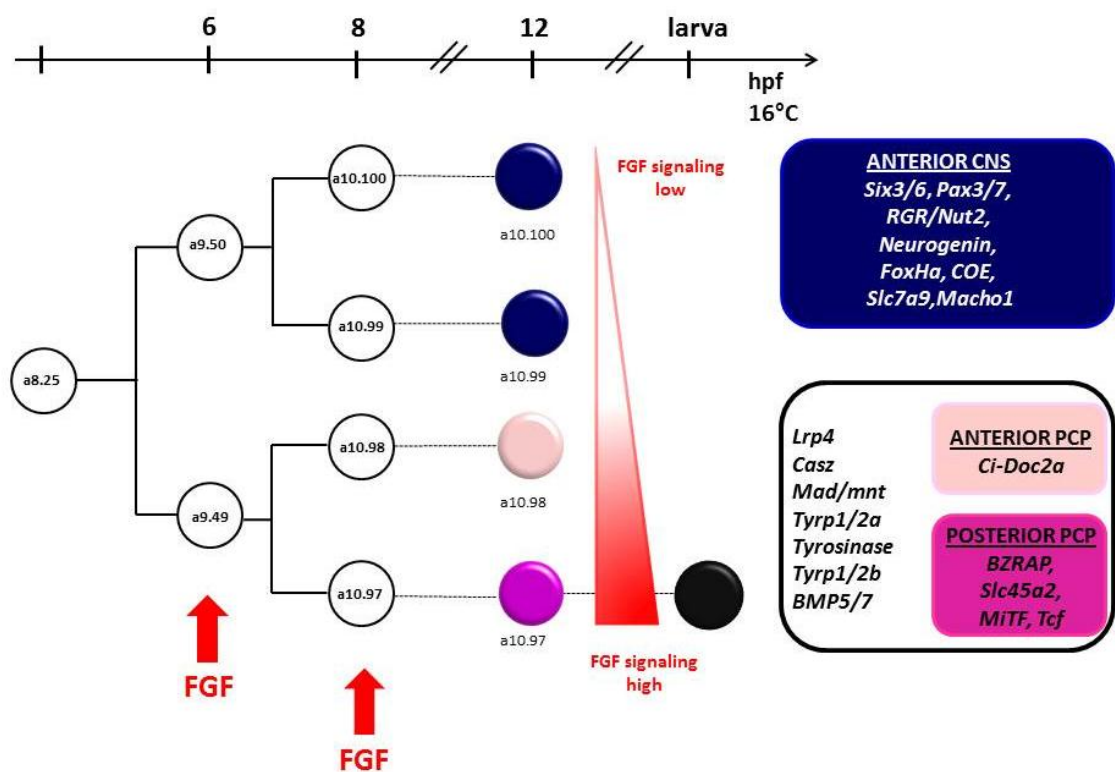
Fig. 4.1 Hypothetic summary model: FGF signaling is required for PCPs induction and CNS anterior-posterior patterning. Cell-lineage scheme of PCPs formation at 16°C, starting from a8.25 cell pair division up to pigmented cell at larval stage. See discussion for a step-by-step description of the model. In blue are shown precursors of Anterior CNS, that arise from a9.50 cell division; in light pink are shown the a10.98s and in pink the a10.97s. in black is represent the pigmented cell that form sensory organs at larval stage. The red arrows indicate FGf induction; the diverse FGF responsiveness of PCPs and anterior CNS precursors is

evidenced by the red gradient.in the blue rectangle are evidence genes expressed in the anterior CNS precursors, in the black rectangle circular line genes expressed in all the four PCPs, in the light pink rectangle genes specifically expressed in a10.98s in the pink rectangle genes expressed just in a10.97s.

Among genes that are expressed in the anterior *Ciona* CNS and show differential expression, it resulted interesting the presence of *slc7a9*, the ortholog of Vertebrate Large Neutral amino acids transporters Small subunit L type amino acid transporter that is responsible for Cystinuria, a human inherited renal and intestinal disease characterized by defective amino acids reabsorption; recently mutation is in this gene is associated also to mental retardation, difficulty in walking and ataxia (Lee et al., 2010). In *Ciona*, *slc7a9* is expressed restrictively in the anterior CNS and from microarray analysis, it shows a robust differential expression when dnFGFR or EtsVp16 are compared to control samples at both 8 and 12 hpf, demarking a role of FGF signaling cascade in controlling this gene expression. Further analyses are necessary to characterize the functional role of *Ci-slc7a9* in *Ciona* CNS differentiation.

4.4 Molecular marker of melanosomal logistic in *Ciona*

The microarray data analysis also permitted to identify new specific markers of *Ciona* pigmented cell lineage (*Ci-Lrp4, Ci-Doc2a, Ci-Bzrap1, Ci-Casz, Ci-Mad/mnt, ,Ci-SLC45A2, Ci-Rab38/32/rab-rp1/ltd*) and to select several genes that may function as key factors to prime the *Ciona* melanosomal logistic, including biogenesis and transport; among these genes, none have been characterized for melanosomal logistic functions in *Ciona* and they represent new FGF candidate targets involved in pigmented cell formation. Even if in literature there are several examples of these genes involvement in pigmented organelle formation (Ohbayashi and Fukuda, 2012; Raposo and Marks, 2007) no evidences are reported about the role that FGF signaling may have in the regulation of these genes expression and more generally on the melanosomal trafficking. For examples the microarray analysis I performed allowed me to identify the Rab GTPase family member, *Ci-Rab38/32/rab-rp1/ltd*, as a new specific marker genes expressed in *Ciona* pigmented cell lineage from neurula to tailbud stage; considering that I identified Ets Binding sites in its promoter region, required for proper spatio-temporal espression of Rab32-38 in pigment cell lineage, we can suppose that FGF signaling cascade direcltly control its expression. Rab proteins are small GTPases considered as central regulators of protein trafficking. They recruit molecular motors to organelles and transport-vesicles, coordinate intracellular signalling with membrane trafficking and play a critical role in the definition of organelle identity (Grosshans et al., 2006). These proteins constitute the largest family of small Ras-like GTPases with 11 identified in yeast and more than 60 members in humans, with 44

subfamilies performing specific sets of functions at distinct subcellular localisations and tissues. Among these, Rab32 and Rab38 are two highly homologous small Rab GTPase specifically expressed in Vertebrate melanocytes. They act in a functionally redundant way regulating skin melanocyte pigmentation. A recessive Gly-19-to-Val mutation in Rab38 have been identified in the hypopigmented *cht* mutant mice, although loss of functional Rab38 causes a mild defect in coat color (Loftus, 2002). Depletion of the closely related Rab32 in *cht* melanocytes causes a dramatic loss of pigmentation, indicating that the near normal pigmentation in *cht* melanocytes results from functional compensation by Rab32 (Wasmeier et al., 2006). In melanocytes lacking Rab38 and Rab32, tyrosinase is degraded after exit from the trans-Golgi network (TGN), indicating that Rab38 and Rab32 regulate a step in the transport of melanogenic enzymes from the TGN to immature melanosomes (Wasmeier et al., 2006). An important feature of the Rab family is that Rab orthologs tend to perform similar functions even in divergent taxa. Functional redundancy between mammalian Rab38 and Rab32 is consistent with the presence of only a single homologue in other species, such as Rab-RP1 in *D.melanogaster* (Fujikawa et al., 2002b). Rab-RP1 is mutated in *lightoid*, a *D. melanogaster* eye color phenotype exhibiting defects in pigment granule synthesis (Ma, 2004). In *Ciona*, *Ci-Rab38/32/rab-rp1/ltd* is the only ortholog of Rab32 and Rab38; mutation in Ci-Rab38/32 protein in the GTP binding pocket, the same responsible for *cht* mouse mutation (Rab38/32^{G19V} where the glycine 19 is conserved in >50% of Rabs) leads to larvae lacking one or both pigmented sensory cells and resulting in a faint pigmentation process alteration; these results demonstrated that Rab38 function is conserved also in *Ciona*. Furthermore I prepared a Ci-Rab38/32 double mutant (Rab38/32$^{G19V+T23N}$) with one mutation at Gly 19 like in *cht* mouse and another one in Thr 23, as reported for a GTP binding mutant for human Rab32 (Alto, 2002). The targeted expression of Rab38/32$^{G19V+T23N}$ in *Ciona* pigment cell precursors induced a stronger alteration of sensory organs pigmentation, resulting in 95% mutated larvae lacking of one or both pigmented organs (Fig. 3.32). These observations suggest an evolutionarily conserved role for the Rab38/Rab32-related subgroup of Rab proteins in the biogenesis of specialized lysosome-related organelles (LROs) such as pigmented sensory organs in *C. intestinalis*, eye pigment granules in *D. melanogaster*, and mammalian melanosomes.

For a better understanding of Rab38/Rab32 function and the trafficking pathways regulated by this subfamily of Rab proteins in melanocytes and other cell types, the identification of interacting partners for Rab38 and Rab32 will be of great interest. Even though the process leading to pigmented cell formation in *Ciona* is poorly characterized in comparison to the melanosomal biogenesis of Vertebrate melanocytes, studies in *Ciona* could fill an important niche in delineating specific aspects that are more complex to delineate in

Vertebrate melanocytes. *Ciona* pigmented cells provide a good opportunity to analyse melanosome biogenesis in *vivo* during the short time window of intense melanogenesis. Most of studies have been carried out on cultured skin melanocytes cell lines that have been extensively used to analyse pathways leading to mature melanosome biogenesis. Unlike skin melanocytes, *Ciona* pigmented sensory organs retain their melanosomes after synthesis, making it possible to gain novel insight into the regulation of melanosome biogenesis and stability.

Furthermore *C. intestinalis* has a small, compacted genome (~160 Mb) that is less redundant than those of vertebrates (Satou et al., 2003a) which is probably related to the twofold to threefold duplication of genomes during Vertebrate evolution (Donoghue and Purnell, 2005). Only 25 genes referred to Rab GTPase family members are annotated on ANISEED database and this peculiarity represents a good advantage to define the sub-compartmentalization of the endosomal system in a highly specialized cells as pigmented cells and to identify protein–protein interactions that mediate melanosome formation, avoiding the redundancy typical of Vertebrates.

Finally another open question on melanosome biogenesis is how the solute gradients regulates protein sorting. Transporters are becoming increasingly recognized as key regulators of membrane dynamics in the endocytic pathway (for example, in cholesterol metabolism (Maxfield and Mondal, 2006). From microarray data analysis I found two interesting genes that are differentially expressed when FGF signaling is pertubated at both 8 and 12 hpf: *Ci-LRP4* and *Ci-slc45a2*, a low density lipoprotein-receptor and a solute carrier transporter family member respectively. The characterization of melanosome membrane transporters in a simpler endosomal membrane compartments as in *Ciona* pigmented cells provides an unprecedented opportunity to dissect how transporters modulate membrane dynamics *via* the lumenal environment of source and target compartments. Thus, this dark organelle provides a good opportunity to shed light on the future of endosomal trafficking and the formation of specialized organelles, such as LROs.

APPENDIX I

Quality Control (QC) analysis performed on microarray data set includes:

Box whisker plot of normalized microarray data

In order to visualize sample data distribution after normalization, I applied for a Box Whisker Plot; the key features of a box plot are the line in the middle of the box, representing the median, that is the 50th percentile; the either end of the box represents the 25th percentile (bottom edge of the box), the 75th percentile (top edge of the box), 95th percentile (top edge of vertical line) and 5th percentile (bottom edge of vertical line). The whiskers are horizontal lines connected to the box indicating the largest and smallest values not considered outliers. Outliers, represented by small horizontal lines, are values that lie more than 1.5 times the interquartile range from the first of third quartile (the edges of the box).

Box whisker plot is a useful way to quickly compare the probe intensity levels between the arrays of a dataset; if one or more arrays have intensity levels which are drastically different from the rest of the arrays, it can be immediately visualized. For microarray data, these graphs are always constructed using log2 transformed probe intensity values, as the graph would be virtually unreadable using raw values.

After the normalization, all the median lines for each sample are centered to the baseline, represented by the linear value = 1 or $\log_2$ value = 0; in the Figure A.1, the base-2 logarithm of the normalized intensities of the 24 samples are shown with the black lines representing the median values all lined at y = 0.

Boxes represent data distribution, where either end of the box represents the upper and lower quartile; as reported in the Figure A.1, all the 24 boxes show a symmetry in data distribution and they are all centered over the median line at each contain the 25% of genes above and below the median; the whiskers encompass genes within 1.5 deviations of the interquartile range while the few remaining transcripts, represented by individual dots, extend vertically above and below the whiskers between the $\log_2$ intensity values of +5 and -5. Applying these plots, I demonstrated that all 24 microarray chips show an equal data distribution that allow me to compare them each other and proceed with the statistical analysis.
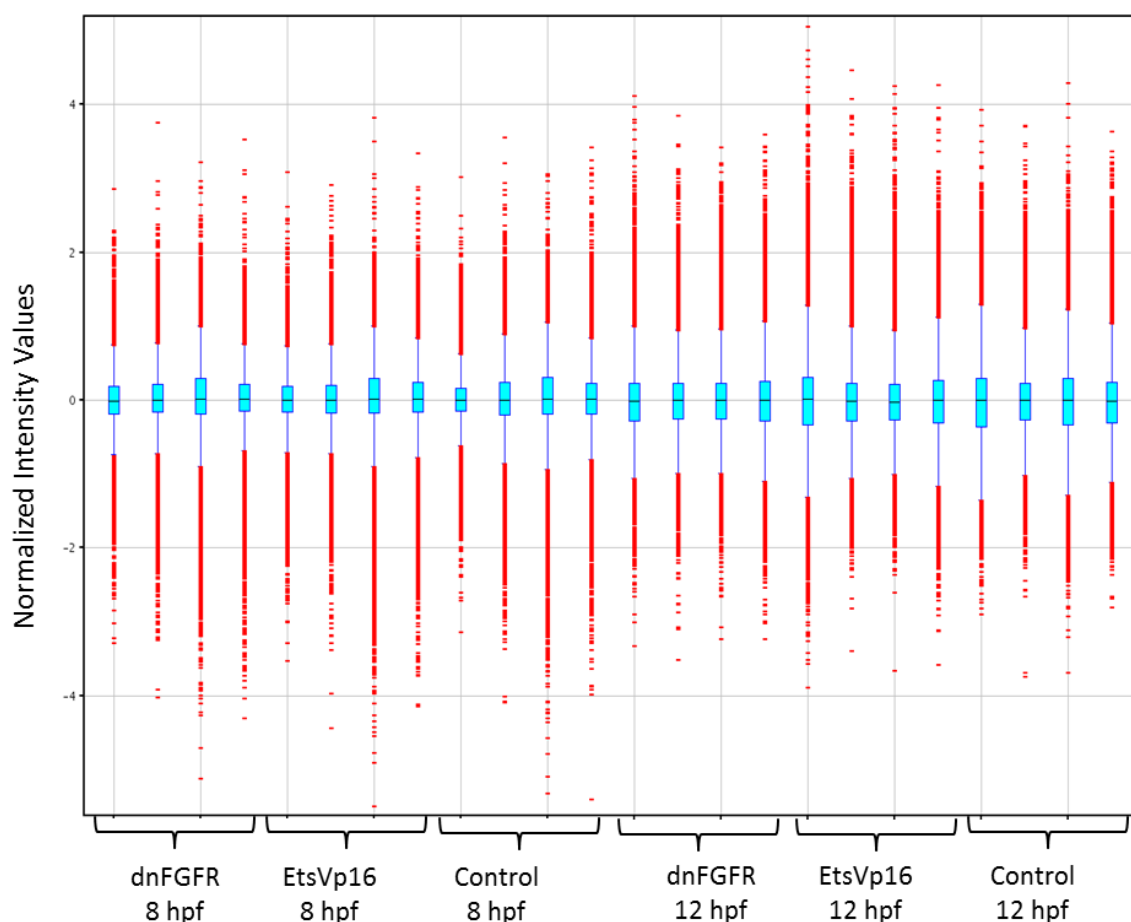
Fig. A.1 Box whisker plot of normalized microarray data set. In the plot, the blue boxes represent data distribution,while the black lines in the middle of the box, the median, that is the 50th percentile; the either end of the box represents the 25th percentile (bottom edge of the box), the 75th percentile (top edge of the box), 95th percentile (top edge of vertical line) and 5th percentile (bottom edge of vertical line). The whiskers are red horizontal lines connected to the box indicating the outliers, represented by small horizontal lines. On the *y*-axis, base-2 logarithm of the normalized intensities values are present; all the four biological replicates for each condition (control, dnFGFR and EtsVp16) and for the two developmental stage (8 and 12 hpf) are shown, all centered over the median line.

Scatter plots for biological replicate comparison

The scatter plot is a two dimensional plot in which a vector is plotted as a point having the coordinates equal to the components of the vector. Each axis corresponds to an experiment and each expression level corresponds to an individual transcripts that is represented as a point. In such a plot, transcripts with similar expression levels will appear somewhere on the first diagonal (the line y = x) of the coordinate system. A transcript that has an expression level very different between the two experiments will appear far from the diagonal.

In order to control how similar biological replicates are, I created scatter plots to compare transcript expression level data between two replicates from Control, dnFGFR and EtsVp16 condition respectively at both 8 and 12hpf; to create scatter plots, I selected the

log$_2$ expression values of two out the four biological replicates isolated for each conditions. As reported in the Figure A.2 , the two boundaries represent a +/- two fold change in transcript expression values; each transcript is represented as a single dot colored upon the scale range representing the expression values from blue, referred to a decreased expression to red, representing an increased expression; yellow dots are transcripts with an expression level close to zero.

As shown in the Figure A.2, most of the transcripts are present within the +/- two fold limit and they are lined along the diagonal of the scatter plot, while very few dots are far from the diagonal demonstrating that between biological replicates just some transcripts have different expression level.

These analysis confirmed that the replicates isolated for the three conditions tested have a comparable expression for the same transcripts analysed, demonstrating that there is a biological similarity among replicates.
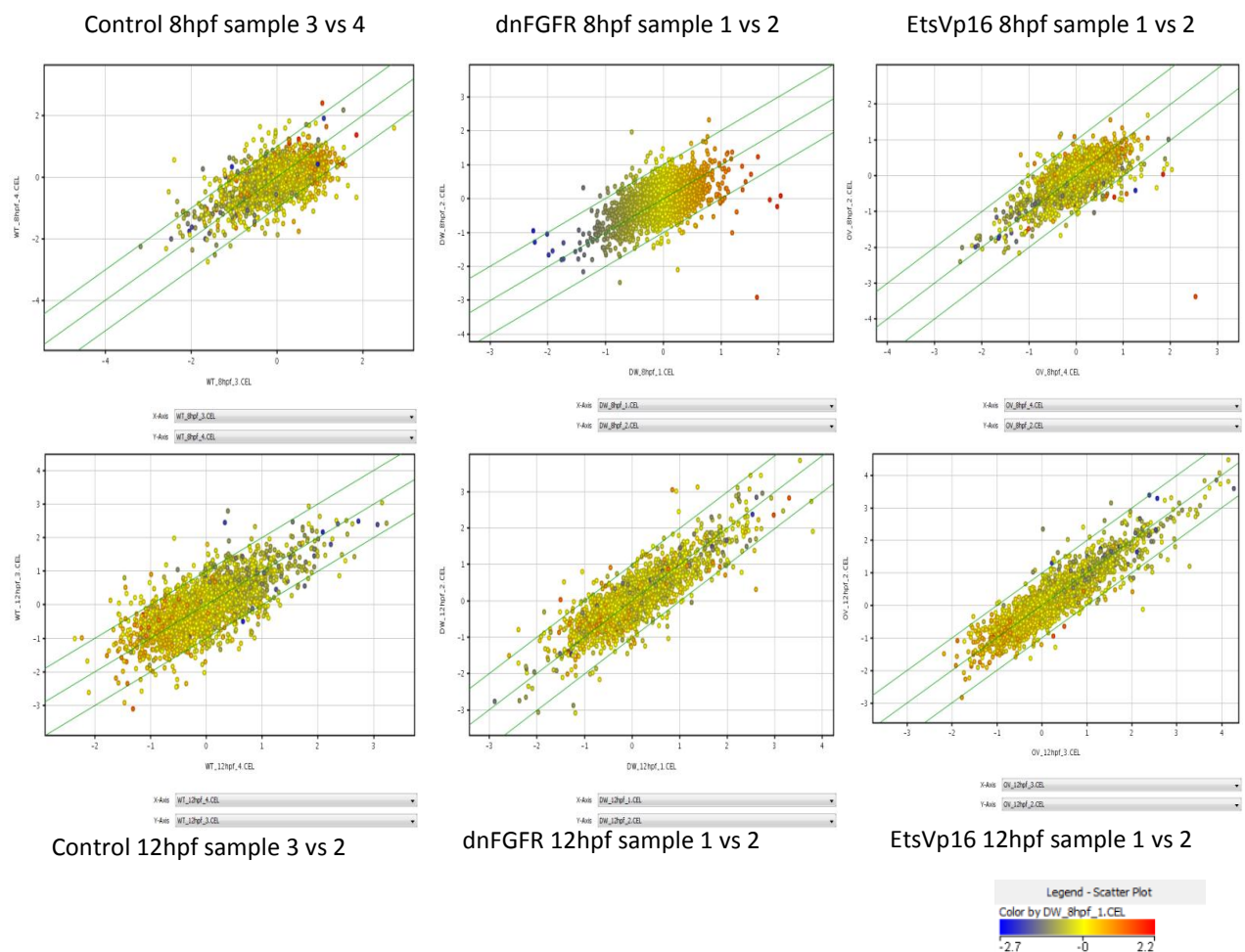
Fig. A.2 Scatter plot for biological replicates. Plots are referred to the comparison transcript expression data of biological replicates among each others; each axis of the plot corresponds to an experiment and each expression level corresponds to an individual transcripts, represented as a colored point associated to its expression (see legend-colour range). The green lines are referred as +/- two fold limits.

Principal Component Analysis (PCA)

This method is used to reduce multidimensional datasets to lower dimensions for analysis and it can determine the key features of high-dimensional datasets. In the context of microarray analysis, PCA essentially clusters arrays in groups calculating the PCA scores and visually represents them in a 3D scatter plot; the scores are used to check data quality. In the Figure A.3, the arrays are represented by colored shaped points that allows viewing the separations between groups of replicates. Ideally, replicates within a group should cluster together and separately from arrays in other groups. The PCA components, represented in the X, Y and Z axes are numbered 1, 2, 3... according to their decreasing significance.

I applied to 3D-PCA plot to visualize in space all the 24 microarray chips; I observed two main groups: one contained all 12 samples collected at 8hpf, another one with the 12 samples collected at 12 hpf; this clustering demonstrates that the replicates for the same time point group together in space demonstrating that they are biologically similar.
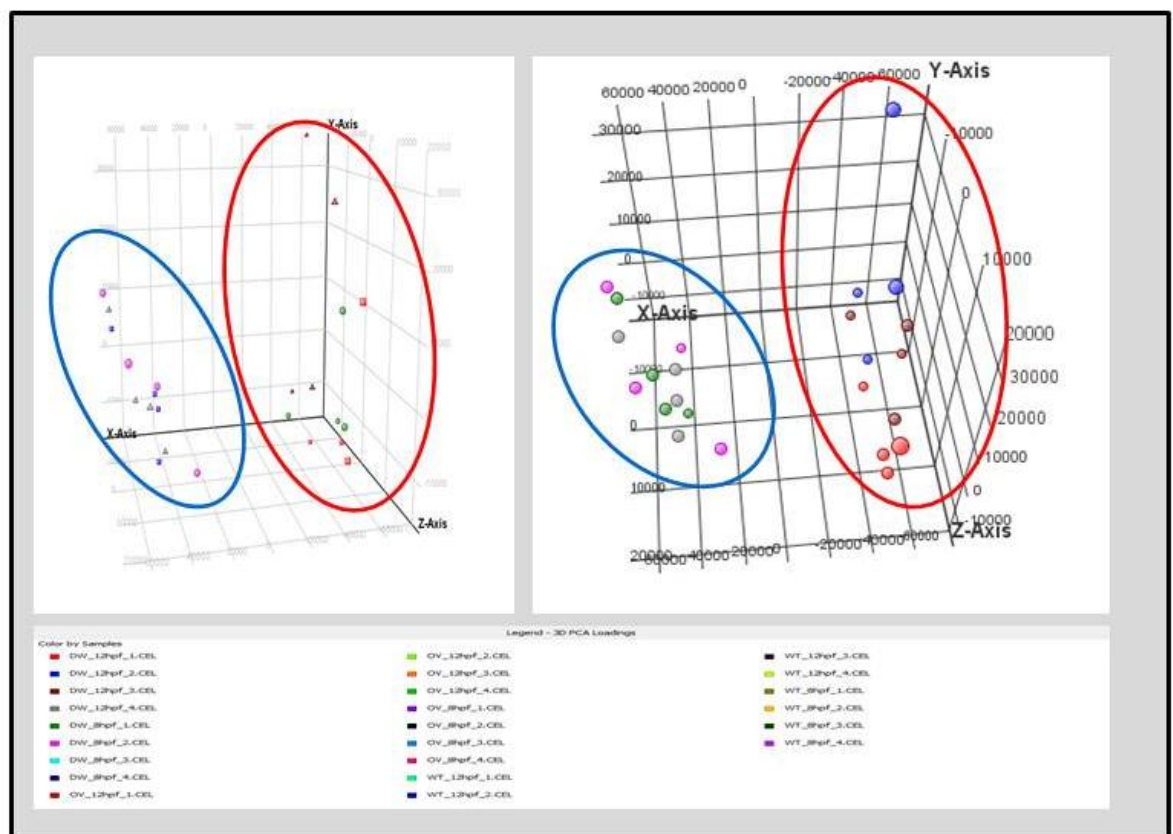
Fig. A.3 3D-Principal Cluster Analysis (PCA). PCA analysis clusters arrays in groups calculating the PCA scores and visually represents them in a 3D scatter plot; the arrays are represented by colored shaped points (see legend) that allows viewing the separations between groups of replicates. Ideally, replicates within a group should cluster together and separately from arrays in other groups. The PCA components, represented in the X, Y and Z axes are numbered 1, 2, 3... according to their decreasing significance.

Correlation Analysis

The Correlation Plot shows the correlation analysis across arrays that finds the correlation coefficient for each pair of arrays using Pearson Correlation Coefficient. Pearson Correlation calculates the mean of all elements in vector *a*; then it subtracts that value from each element in *a* and calls the resulting vector A. It does the same for *b* to make a vector B. Correlation Coefficient (*r*) is then computed as A*B/(|A||B|).

I applied for a correlation plot to visualize the correlation coefficients calculated for all the 24 microarray chips; the analysis revealed high correlation between and within each group as expressed in the Pearson's correlation coefficient *r*, that ranged from 0.877 to 0.998 with an average r of 0.963 within the several arrays. The Correlation Analysis is displayed as a heat map to quickly visualize the result; the colour scale range on the right shows the correlation of samples and runs from black (*r* = 0.75), reflecting medium correlation, to red (*r* = 1) , reflecting high correlation. As shown in Figure A.4, all four biological triplicates for a specific condition have correlation coefficients *r* very high, with an average of 0.96, resulting in a cluster of red squares, while comparing data between arrays at 8 and 12 hpf, the correlation coefficients are lower resulting in a cluster of black squares.
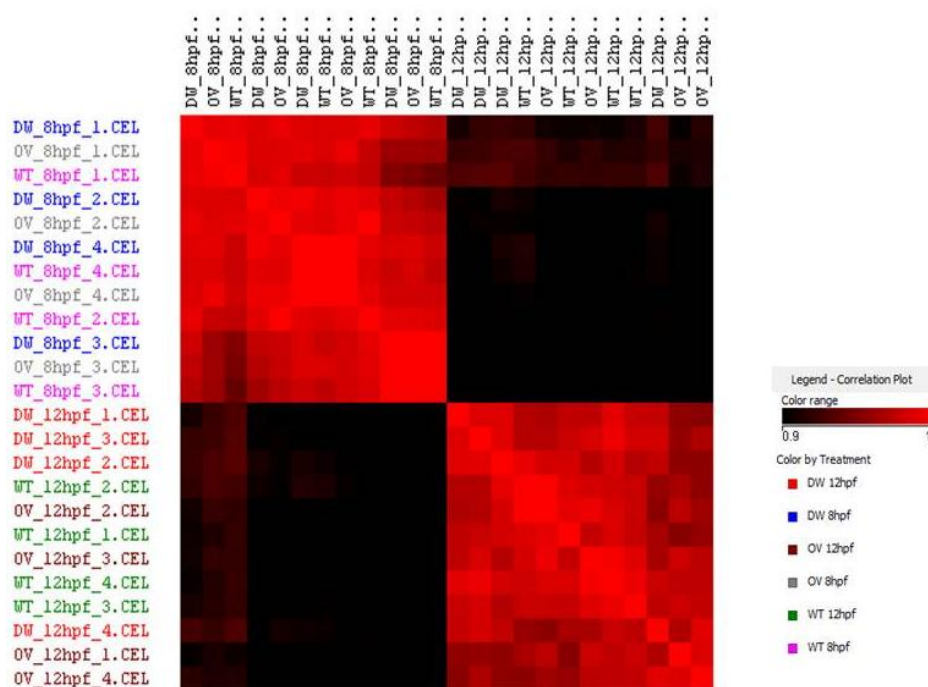
Fig. A.4 Heat map displaying the Correlation Analysis. Pearson Correlation Coefficient have been calculated for each pair of arrays for all the 24 chips that are on both *x*- and *y*-axis (see legend for color sample association); the colour scale range on the right shows the correlation of samples, from black, r = 0.75, to red, r = 1.

The Affymetrix hybridization controls

This is a view to depict the hybridization quality. Hybridization controls are composed of a mixture of biotin-labelled cRNA transcripts of bioB, bioC, bioD, and cre prepared in staggered concentrations (1.5, 5, 25, and 100pm respectively). This mixture is present in the Affymetrix Hybridization mix and is spiked-in into the hybridization cocktail; bioB is at the level of assay sensitivity and should be called Present at least 50% the time; bioC, bioD and cre must be present all of the time and should appear in increasing concentrations. The *x*-axis in this graph represents the controls and the *y*-axis, the log of the Normalized Signal Values. When I applied for this hybridization control view, I noticed that 3 out 24 samples have a different pattern in comparison to the others (Fig. A.5). In order to avoid biasness in results, I performed two statistical analyses, one including them and another excluding this 3 samples. In both case, I obtained almost a similar result and I continue analysing data excluding them (Fig. A.6).
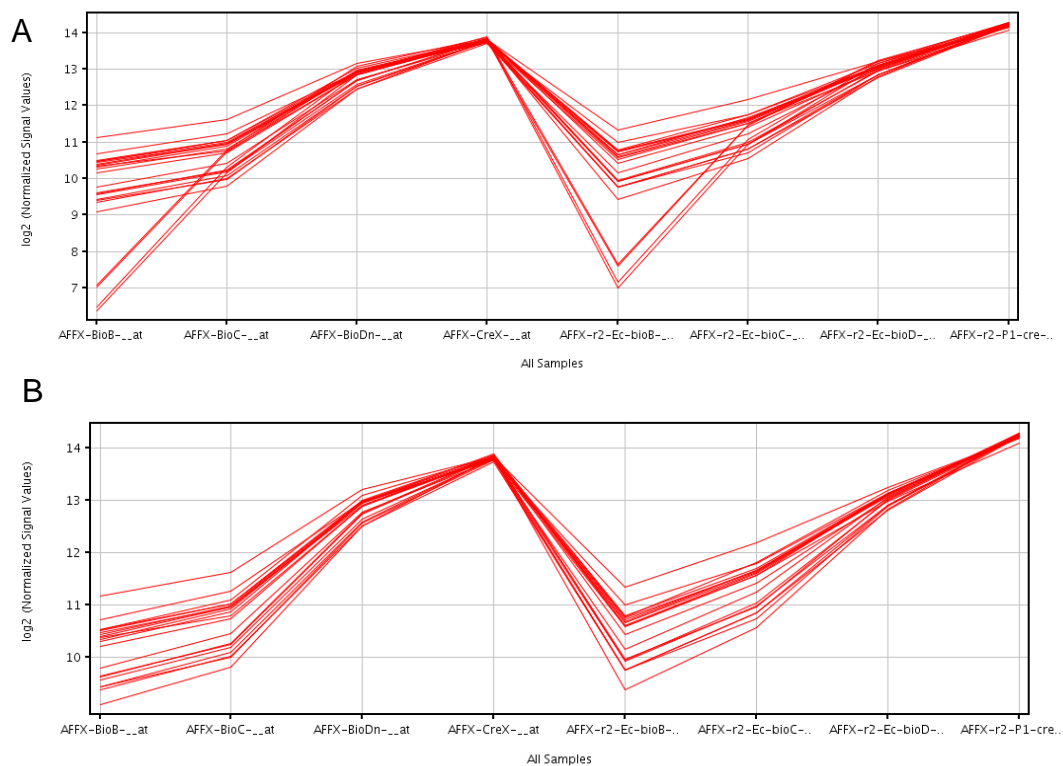
Fig. A.6 The hybridization controls expression patterning (A,B). Hybridization controls are biotin-labelled cRNA transcripts of bioB, bioC, bioD, and cre prepared in staggered concentrations (1.5, 5, 25, and 100pm respectively) and present on chips as positive controls; in the graph, the *x*-axis represents the controls and the *y*-axis, the $\log_2$ of the Normalized Signal Values. Three samples show an ambiguous expression patter (A) and they have been excluded for further analysis (B).

Quality control on Probes: Filter Probeset by Expression Values

Raw data values can be filtered based on their signal intensity values. This step is useful to reduce the false positive genes in the results removing very low signal values or those that have reached saturation on the chips. Filtering is done on raw data with percentile cut-off: upper cut-off is 100th percentile while lower cut-off is 20th percentile; in this way I'm removing very low signal values, lower percentile cut-off 20% or those that have reached saturation, lowering the upper percentile cut-off from 100%.

In this operation the lowest 20 percentile of all the intensity values are removed and a new profile plot of filtered entities is generated. This operation is performed on the raw signal values while the plot is generated using the normalized (not raw) signal values (Fig. A.7). Starting from 30969 transcripts, applying this filtering method I selected 24568 transcripts that I considered to be good to proceed further in the analysis.
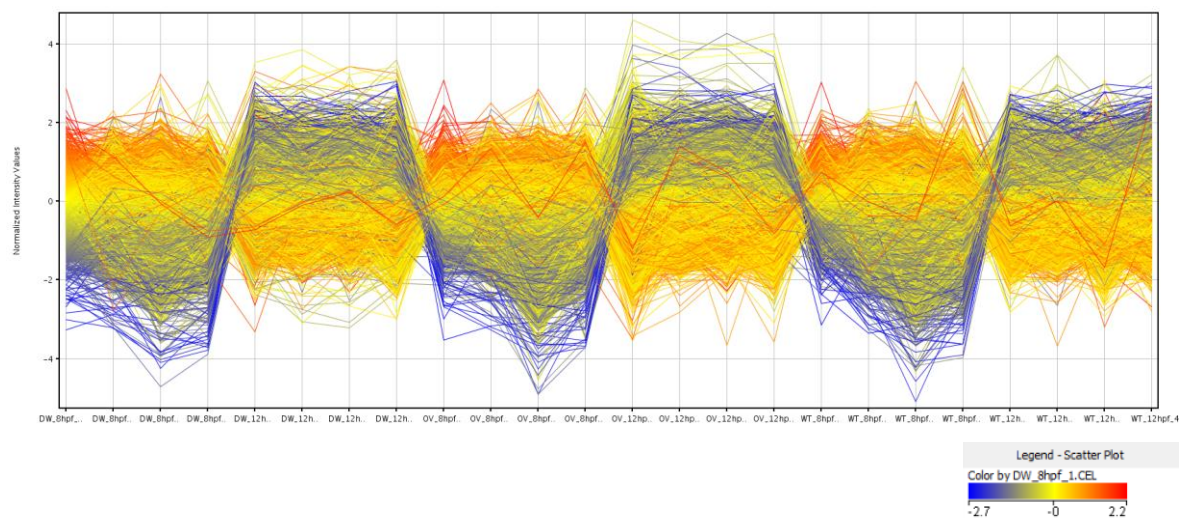
Fig. A.7 Filter Probesets by Expression values. Raw data values have been filtered based on their signal intensity values with an upper cut-off is 100th percentile and a lower cut-off of 20th percentile; in the graph, the x-axis represents the expression values for all the 24 samples and the y-axis, the $\log_2$ of the Normalized Signal Values.

APPENDIX II

The ribonucleic probes listed selected from microarray dataset for the synthesis of antisense ribonucleic probes starting from cDNA clones present in N. Satoh *C. intestinalis* gene collection 1 and Gateway library, available in the laboratory.

Genes selected from microarray dataset for ribonucleic probes synthesis.	
Gene's name	**Corresponding clone in N. Satoh gene collection 1 (R1CiGC) or Gateway libray (VES)**
WFIKKN1	ciad35c03 (R1CiGC04a21)
Ci-Six3/6	cicl21e08 (R1CiGC11m13)
βγ CRYSTALLIN	cilv10f04 (R1CiGC25f12)
Sushi/Ci-complement recptor-1B	cign004j14 (R1CiGC41f02)
GNRHR	cilv11e18 (R1CiGC25l12)
Ci-msxb	cign067l18 (R1CiGC42h24)
Pax3/7	cign035n06 (R1CiGC42e20)
DGKI	cibd026k14 (R1CiGC37n12)
KREMEN1	cilv20p01 (R1CiGC26h10)
EBF_COE	ciad13f14 (R1CiGC02i14)
frizzled1/2/7 putative / Frizzled receptor	cieg35h18 (R1CiGC21o13)
fzd1 / fzd2 / fzd3 / fzd6 / fzd7 / frizzled receptor / fz	cits38k23 (R1CiGC34l06)
EFNA1	cits38k23 (R1CiGC34l06)
Pax6	cieg37l08 (R1CiGC21d13)
DHH,hedgehog homolog 1	cien111024 (VES87_B20)
frizzled4	cieg11m06 (R1CiGC17m02)
FoxH-a,Foxh1	cies002j17 (R1CiGC41a17)
KLH20	citb38e24 (R1CiGC32f03)
SLC24A3	ciad60d06 (R1CiGC06m23)
Ci-macho1	cicl14k05 (R1CiGC10a23)
RABEPK	cieg16n12 (R1CiGC18g07)
Groucho	cicl21a01 (R1CiGC11k07)
EVI1	cilv049l02 (R1CiGC44e10)
Btsz	citb6p3 (R1CiGC29l23)
Bcmo1	ciad57a22 (R1CiGC05j16)
Aebp2	citb6g18 (R1CiGC29h11)

Ci-Tyrp1/2a	cieg01h02 (R1CiGC15f05)
RLBP1/CRALBP	citb30d10 (R1CiGC31h05)
Blk	cilv21j03 (R1CiGC26j04)
Ci-Tyrp1/2b	citb19d22 (R1CiGC31o17)
ATPAF1	cilv069a04 (R1CiGC44b23)
OdsH	citb6p3 (R1CiGC29l23
RAB3GAP1	citb049b24 (R1CiGC45d02)
Ci-Tyrosinase	ibd043b13 (R1CiGC38o19)
Ci-Noto2	citb41l04 (R1CiGC33c19)
DEAF1	ciad11l19 (R1CiGC02m05)
CENTD1,RhoGAP15B	citb11c24 (R1CiGC30a06)
ARRDC2	cilv11b02 (R1CiGC25j20)
CiGl	cien153938 (VES94_I02)
FoxD3	ciad01a24 (R1CiGC01a05)
RHBDL1	citb8o13 (R1CiGC29n08)
Ci-Hunchbacklike,CASZ1	cima869137 (VES66_L10)
TRPA1	citb16l18 (R1CiGC31a11)
Rfxdc1	cieg082o22 (R1CiGC40m20)
GJA10	cibd062b06 (R1CiGC38j11)
Slc43a3	cinc011h13 (R1CiGC45a07)
Pip5k3	cien196402 (VES100_H18)
Tcf-Lef	cign010o13 (R1CiGC41l16)
BMP5	cicl49a12 (R1CiGC13p03)
GSC	citb068b12 (R1CiGC46m03)
Bzrap1	ciad56k04 (R1CiGC05h08)
aub	ciad64j14 (R1CiGC06d17)
Mitf	cieg52a12 (R1CiGC23f14)
Fringe	cilv41b12 (R1CiGC28k08)
ELK3,Elk1	ciad088b10 (R1CiGC37e01)
Rfxdc1	cicl06n02 (R1CiGC09c11)
TRPA1	cibd062b06 (R1CiGC38j11)
ARRDC2	cieg082o22 (R1CiGC40m20)
pcs,Sh3bp5	cien153938 (VES94_I02)
Oxysterol binding protein-related protein 1	cicl53n23 (R1CiGC14g18)
Dusp6.9	ciad37c04 (R1CiGC04m07)

LRP4,Ci-Low-density lipoprotein receptor-related gene 1	cign049p19 (R1CiGC42f11)
DEAF1	citb38a22 (R1CiGC32b23)
btsz	citb11c24 (R1CiGC30a06)
RABEPK	ciad57a22 (R1CiGC05j16)
DGKI	cicl21a01 (R1CiGC11k07)
GSC	cibd026k14 (R1CiGC37n12)
Ascl2,Achaete-Scute a-like2	ciad56k04 (R1CiGC05h08)
SLC45A2	citb32n03 (R1CiGC31j20)
ARNT	cicl17l13 (R1CiGC10h11)
MMR/ Macrophage mannose receptor 1 precursor	cien41888 (VES76_H24)

APPENDIX III

The most represented Gene Ontology classes recovered from GO analysis on differential expressed transcript list obtained comparing dnFGFR to Control at 12 hpf.

GO analysis on transcripts differential expressed when comparing dnFGFR to control at 12hpf					
GO.id	# transcript selected	p.sel	p.value	adj.p	definition
GO:0043473	16	1.670146	7.00E-17	1.14E-13	pigmentation
GO:0030318	8	0.835073	5.12E-11	8.34E-08	melanocyte differentiation
GO:0004857	12	1.25261	3.61E-09	5.88E-06	enzyme inhibitor activity
GO:0006345	4	0.417537	2.33E-08	3.79E-05	loss of chromatin silencing
GO:0035059	4	0.417537	2.33E-08	3.79E-05	RCAF complex
GO:0030156	4	0.417537	2.33E-08	3.79E-05	benzodiazepine receptor binding
GO:0014034	6	0.626305	7.28E-08	0.000118	neural crest cell fate commitment
GO:0007500	5	0.521921	1.71E-07	0.000277	mesodermal cell fate determination
GO:0030050	4	0.417537	1.88E-07	0.000306	vesicle transport along actin filament
GO:0048570	7	0.730689	2.49E-07	0.000404	notochord morphogenesis
GO:0004806	6	0.626305	4.59E-07	0.000743	triglyceride lipase activity
GO:0031261	4	0.417537	1.01E-06	0.001628	DNA replication preinitiation complex
GO:0046513	7	0.730689	1.64E-06	0.002653	ceramide biosynthetic process
GO:0007164	6	0.626305	2.10E-06	0.003392	establishment of tissue polarity
GO:0051270	6	0.626305	2.10E-06	0.003392	regulation of cellular component movement
GO:0008191	8	0.835073	2.57E-06	0.004143	metalloendopeptidase inhibitor activity
GO:0048319	3	0.313152	3.71E-06	0.005983	axial mesoderm morphogenesis
GO:0046022	3	0.313152	3.71E-06	0.005983	positive regulation of transcription from RNA polymerase II promoter during mitosis
GO:0048625	3	0.313152	3.71E-06	0.005983	myoblast cell fate commitment
GO:0000095	3	0.313152	3.71E-06	0.005983	S-adenosylmethionine transmembrane transporter activity
GO:0015805	3	0.313152	3.71E-06	0.005983	S-adenosylmethionine transport
GO:0031489	4	0.417537	3.97E-06	0.006383	myosin V binding
GO:0030261	10	1.043841	4.70E-06	0.007553	chromosome condensation
GO:0001736	11	1.148225	6.08E-06	0.009771	establishment of planar polarity
GO:0048557	4	0.417537	1.25E-05	0.020031	embryonic digestive tract morphogenesis
GO:0000916	4	0.417537	1.25E-05	0.020031	contractile ring contraction involved in cell cycle cytokinesis
GO:0048619	4	0.417537	1.25E-05	0.020031	embryonic hindgut morphogenesis
GO:0030178	15	1.565762	1.33E-05	0.021321	negative regulation of Wnt receptor signaling pathway
GO:0000285	3	0.313152	2.37E-05	0.037902	1-phosphatidylinositol-3-phosphate 5-kinase activity

GO:0046849	3	0.313152	2.37E-05	0.037902	bone remodeling
GO:0021903	3	0.313152	2.37E-05	0.037902	rostrocaudal neural tube patterning
GO:0004771	3	0.313152	2.37E-05	0.037902	sterol esterase activity

APPENDIX IV

Results for the neighbors of level one analysis of *Ci-Tyrosinase*, *Ci-Tcf/LEF1*, *Ci-Tyrp1/2a*, *Ci-Tyrp1/2b* and *Ci-Rab38/32* from transcriptional regulatory network.

Table 1: Neighbors of level one of *Ci-Tyrosinase*		
Probe Set IDs	SCC	*Ciona* name
MKG.13.94.1_at	0.94	Tyrp1/2a
MKG.26.42.1_at	0.9374	SLC45A2
MKG.123.32.s0_s_at	0.9322	Bmp2
MKG.33.7.1_at	0.9157	Cas,BC035954
MKG.132.36.s0_at	0.9122	Copper-transporting ATPase 1
MKG.118.37.1_at	0.9113	Zinc finger protein 366
MKG.42.3.1_at	0.907	Tyrp1/2b
MKG.557.6.1_at	0.9026	Nahoda
MKG.36.63.s0_at	0.9017	ATP-dependent DNA helicase II
MKG.245.12.1_at	0.9	Carbohydrate sulfotransferase 11
MKG.43.49.s0_at	0.8965	RAB3GAP1
MKG.51.66.s0_at	0.8957	CD209
MKG.422.15.s0_at	0.893	Zinc finger protein mex-6
MKG.26.68.1_at	0.8887	TAGLN
MKG.1.92.1_at	0.887	ATP6V1B1
MKG.110.2.1_s_at	0.8791	Nicotinic acetylcholine receptor subunit alpha 9-II
MKG.422.8.1_at	0.8774	Pip5k3
MKG.28.60.1_at	0.8765	Ephrin type-B receptor 1A precursor
MKG.36.62.1_at	0.8678	LRP2 like
MKG.118.34.s1_at	0.867	KIAA0297
MKG.1106.3.1_at	0.8643	Phosphoglyceromutase
MKG.113.2.1_at	0.8635	S-adenosylmethionine-dependent methyltransferase activity
MKG.261.7.1_at	0.8617	Chromosomal replication initiator protein dnaA
MKG.158.28.1_s_at	0.8609	UGP
MKG.26.70.1_at	0.8591	Chromosomal replication initiator protein dnaA
MKG.13.65.s0_s_at	0.8574	CYP26A1
MKG.33.5.s0_s_at	0.8574	Ci-Hunchbacklike,CASZ1
MKG.302.11.1_at	0.8548	Xylose isomerase 2
MKG.330.7.1_at	0.8548	SDCBP
MKG.110.2.1_at	0.8539	Neuronal acetylcholine receptor protein
MKG.1464.2.1_at	0.8504	Hypothetical sodium-dependent transporter
MKG.97.38.s0_at	0.8504	Intelectin 1b precursor
MKG.118.34.s0_at	0.847	Hypothetical protein MG098
MKG.422.12.1_s_at	0.847	Pip5k3
MKG.680.6.s0_s_at	0.8461	UGP
MKG.516.3.1_at	0.8452	Gal3st3
MKG.113.15.1_s_at	0.8435	Tetraspanin-9
MKG.260.4.s0_s_at	0.8409	AP3D1
MKG.4.104.1_at	0.8409	Ribonuclease Z

MKG.20.69.s1_s_at	0.84	FCN1
MKG.92.13.s0_at	0.8383	CHMP5
MKG.1425.1.1_s_at	0.8374	eIF-4-gamma 3
MKG.330.7.1_s_at	0.8357	SDCBP

Table 2: Neighbors of level one of *Ci-Tcf/ELF1*		
Probe Set IDs	SCC	*Ciona* name
MKG.61.28.1_at	0.9157	LEF1
MKG.1464.2.1_at	0.8861	Xylose isomerase 2
MKG.557.7.1_at	0.8861	ATP synthase beta chain
MKG.422.12.s1_s_at	0.8843	PIP5K3
MKG.101.41.1_s_at	0.8643	Goliath protein precursor
MKG.113.15.1_s_at	0.8574	Tetraspanin-9
TG0514.261.r.16_at	0.8557	Bmp2
MKG.422.15.s0_at	0.8496	Zinc finger protein mex-6
MKG.20.70.1_at	0.8478	FCN1
MKG.557.6.1_at	0.8435	Nahoda
MKG.392.5.1_s_at	0.8409	NPAL2
MKG.475.1.s0_at	0.8374	Kinetochore protein CTF13
MKG.286.12.1_at	0.8357	Dachsous protein precursor
MKG.40.10.s0_at	-0.8383	RGR
MKG.97.9.1_at	-0.8383	ZFAND1
TG0514.7.f.34_at	-0.84	Collagen alpha 3(VI) chain precursor
MKG.108.29.s0_at	-0.847	PRR6
MKG.1116.1.1_s_at	-0.847	EDEM2
MKG.171.25.1_s_at	-0.8687	tRNA-guanine transglycosylase

Table 3: Neighbors of level one of *Ci-Tyrp1/2a*		
Probe Set IDs	SCC	*Ciona* name
MKG.166.29.1_at	0.94	Ci-Tyrosinase,Tyr
MKG.26.42.1_at	0.9339	SLC45A2
MKG.42.3.1_at	0.933	Tyrp1/2b
MKG.123.32.s0_s_at	0.9313	Bmp2
MKG.110.2.1_s_at	0.9191	Neuronal acetylcholine receptor protein
MKG.26.43.s0_s_at	0.9113	beta4GalNAcTA
MKG.1425.1.1_s_at	0.9061	eIF-4G 3
MKG.110.2.1_at	0.9043	Neuronal acetylcholine receptor protein
MKG.557.6.1_at	0.9035	Nahoda
MKG.422.15.s0_at	0.8965	Zinc finger protein mex-6
MKG.51.66.s0_at	0.8965	CD209
MKG.36.63.s0_at	0.8913	TP-dependent DNA helicase II
MKG.26.70.1_at	0.8861	Chromosomal replication initiator protein dnaA
MKG.28.60.1_at	0.8852	Ephrin type-B receptor 1A precursor
MKG.202.15.1_s_at	0.8835	Lipoprotein-associated coagulation inhibitor
MKG.1.92.1_at	0.8826	ATP6V1B1
MKG.92.13.s0_at	0.8809	CHMP5

MKG.42.1.s0_at	0.88	ABHD4
MKG.97.38.s0_at	0.88	Intelectin 1b precursor
MKG.33.5.s0_s_at	0.873	Ci-Hunchbacklike,CASZ1
MKG.751.5.1_s_at	0.873	Putative MAP kinase activating protein C22orf5
MKG.182.15.1_at	0.8722	Bel
MKG.2741.1.1_at	0.8704	beta4GalNAcTA
MKG.69.28.s0_at	0.8696	MXD1,Mnt
MKG.118.37.1_at	0.8687	Zinc finger protein 366
MKG.96.32.s0_s_at	0.8687	Atp6ap1
MKG.1464.2.1_at	0.8652	Tetraspanin-9
MKG.761.5.s0_s_at	0.8635	RpS7
MKG.330.7.1_at	0.8626	SDCBP
MKG.13.65.s0_s_at	0.8617	CYP26A1
MKG.262.6.s0_at	0.8617	Grrp1,5430432M24Rik
MKG.392.5.1_s_at	0.8591	NPAL2
MKG.260.4.s0_s_at	0.8574	AP3D1
MKG.13.66.s0_at	0.8522	CYP26A1
MKG.422.12.s1_s_at	0.8513	PIP5K3
MKG.43.49.s0_at	0.8487	RAB3GAP1
MKG.234.13.1_at	0.8478	Samd7
MKG.253.20.1_at	0.8478	3B4GalNAcTA
MKG.33.7.1_at	0.8461	Cas,BC035954
MKG.234.13.s0_s_at	0.8443	Samd7
MKG.158.28.1_s_at	0.8435	UGP
MKG.330.7.1_s_at	0.8426	SDCBP
MKG.513.8.s0_at	0.8417	Serine palmitoyltransferase 2
MKG.26.68.1_at	0.8409	TAGLN
MKG.586.3.s0_at	0.8409	Protein-export membrane protein secF
MKG.22.81.s0_s_at	0.84	beta4GalNAcTA
MKG.66.19.1_s_at	0.8374	ATP6V1E1
MKG.127.1.1_s_at	-0.8539	Rgs12
MKG.33.43.1_s_at	-0.8565	2410003A14Rik

Table 4: Neighbors of level one of *Ci-Tyrp1/2b*		
Probe set Ids	SCC	*Ciona* Name
MKG.13.94.1_at	0.933	Tyrp1/2a
MKG.166.29.1_at	0.907	Ci-Tyrosinase,Tyr
MKG.33.5.s0_s_at	0.9061	Ci-Hunchbacklike,CASZ2
MKG.180.7.s1_s_at	0.9035	EXOSC7,Clec3a
MKG.666.1.1_s_at	0.9009	ARRDC2
MKG.26.42.1_at	0.8991	SLC45A2
MKG.1.92.1_at	0.893	ATP6V1B1
MKG.123.32.s0_s_at	0.8904	Bmp2
MKG.422.8.1_at	0.8835	Pip5k3
MKG.69.28.s0_at	0.8826	MXD1,Mnt
MKG.42.1.s0_at	0.8783	ABHD4

MKG.51.66.s0_at	0.8774	CD209
MKG.96.32.s0_s_at	0.8765	Atp6ap1
MKG.111.20.s0_at	0.8748	Pla2g4b
MKG.89.16.1_s_at	0.8704	ANKDD1A
MKG.751.5.1_s_at	0.8661	Putative MAP kinase activating protein C22orf5
MKG.600.3.s1_s_at	0.8583	Rfxdc1
TG0514.1256.f.2_at	0.8548	Ca2+-dependent phospholipid scramblase
MKG.28.60.1_at	0.853	Ephrin type-B receptor 1A precursor
MKG.1425.1.1_s_at	0.8522	eIF-4-gamma 3
MKG.20.69.s1_s_at	0.8522	FCN1
MKG.2741.1.1_at	0.8522	beta4GalNAcTA
MKG.33.7.1_at	0.8522	Cas,BC035954
MKG.118.34.s0_at	0.8487	Hypothetical protein MG098
MKG.36.62.1_at	0.847	LRP2 like
TG0514.23.f.2_at	0.8461	Ornithine decarboxylase
MKG.245.12.1_at	0.8443	Carbohydrate sulfotransferase 11
MKG.16.73.s0_at	0.8435	Hypothetical protein yxaL
MKG.110.2.1_s_at	0.8426	Neuronal acetylcholine receptor protein
MKG.422.15.s0_at	0.8426	Zinc finger protein mex-6
MKG.680.6.s0_s_at	0.8409	UGP
MKG.4.104.1_at	0.8391	RNase Z
TG0514.489.f.10_s_at	-0.8409	ADP-ribosylation factor 3
MKG.86.6.1_at	-0.86	Relaxin-like protein

Table 5: Neighbors of level one of *Ci-Rab38/32*		
Probe Set IDs	SCC	*Ciona* name
MKG.13.94.1_at	0.933	Tyrp1/2a
MKG.166.29.1_at	0.907	Ci-Tyrosinase,Tyr
MKG.33.5.s0_s_at	0.9061	Ci-Hunchbacklike,CASZ1
MKG.180.7.s1_s_at	0.9035	EXOSC7,Clec3a
MKG.666.1.1_s_at	0.9009	ARRDC2
MKG.26.42.1_at	0.8991	SLC45A2
MKG.1.92.1_at	0.893	ATP6V1B1
MKG.261.7.1_at	0.8922	Chromosomal replication initiator protein dnaA
MKG.123.32.s0_s_at	0.8904	Bmp2
MKG.422.8.1_at	0.8835	Pip5k3
MKG.69.28.s0_at	0.8826	MXD1,Mnt
MKG.36.63.s0_at	0.8826	ATP-dependent DNA helicase II
MKG.42.1.s0_at	0.8783	ABHD4
MKG.51.66.s0_at	0.8774	CD209
MKG.96.32.s0_s_at	0.8765	Atp6ap1
MKG.111.20.s0_at	0.8748	Pla2g4b
MKG.89.16.1_s_at	0.8704	ANKDD1A
MKG.751.5.1_s_at	0.8661	Putative MAP kinase activating protein
MKG.422.12.1_s_at	0.8635	Pip5k3
MKG.302.11.1_at	0.86	Xylose isomerase 2

MKG.97.38.s0_at	0.86	Intelectin 1b precursor
MKG.600.3.s1_s_at	0.8583	Rfxdc1
MKG.118.34.s1_at	0.8565	Hypothetical protein KIAA0329
TG0514.1256.f.2_at	0.8548	Phospholipid scramblase 1
MKG.28.60.1_at	0.853	Ephrin type-B receptor 1A precursor
MKG.33.7.1_at	0.8522	Cas,BC035954
MKG.1425.1.1_s_at	0.8522	eIF-4-gamma 3
MKG.20.69.s1_s_at	0.8522	FCN1
MKG.2741.1.1_at	0.8522	beta4GalNAcTA
MKG.118.34.s0_at	0.8487	Hypothetical protein MG098
MKG.36.62.1_at	0.847	LRP2 like
TG0514.23.f.2_at	0.8461	Ornithine decarboxylase
MKG.245.12.1_at	0.8443	Chondroitin 4-sulfotransferase 1
MKG.16.73.s0_at	0.8435	Hypothetical protein yxaL
MKG.422.15.s0_at	0.8426	Zinc finger protein mex-6
MKG.110.2.1_s_at	0.8426	Neuronal acetylcholine receptor protein
MKG.680.6.s0_s_at	0.8409	UDPGP 1
MKG.4.104.1_at	0.8391	RNase Z
TG0514.489.f.10_s_at	-0.8409	ADP-ribosylation factor 3
MKG.86.6.1_at	-0.86	Relaxin-like protein SQ10 precursor

APPENDIX V

Candidate genes regulated by FGF/MAPK/ETS signalling, either positively and negatively, have been selected for biological analyses from the gene list displaying significant (p-value < 0.05) and a robust differential expression (fold change > 1.5) when comparing the two conditions, dnFGFR and EtsVp16, to control at 8 and12 hpf

p-value	regulation	fold change ratio	*Ciona* name

Table 1: Genes selected from the comparison dnFGFR and Control samples at 12 hpf

p-value	regulation	fold change ratio	*Ciona* name
5.09E-06	up	10.582218	BAPX1
5.08E-06	up	3.759728	WFIKKN1
8.27E-04	up	3.660176	Ci-Six3/6
3.30E-05	up	3.622585	βγ CRYSTALLIN
5.06E-05	up	3.0055397	Sushi
6.10E-04	up	2.9282756	GNRHR
0.006571028	up	2.335518	Ci-msxb
0.025059156	up	2.1414404	Pax3/7
2.19E-05	up	2.156824	DGKI
0.020309923	up	1.7883859	KREMEN1
0.01731797	up	1.7814424	EBF_COE
0.003618498	up	1.7792797	FZD1,Frizzled receptor
2.64E-05	up	1.7346642	EFNA1
0.037711434	up	1.7130381	Pax6
0.015872275	up	1.6685985	DHH,hedgehog homolog 1
0.030576145	up	1.6213132	frizzled4
0.013111974	up	1.5353143	MLPH-MyRip
4.43E-05	up	1.456985	FoxH-a,Foxh1
3.88E-05	up	1.485688	KLH20
0.03596837	up	1.4387349	SLC24A3
0.007224274	up	1.4268744	Macho1
4.04E-04	up	1.3958962	FoxHa
1.03E-05	up	1.382520	RABEPK
0.016370108	up	1.3524321	Groucho
0.018553661	up	1.2690612	EVI1
2.29E-05	up	1.215621	Btsz
0.023319472	up	1.181123	Bcmo1
0.04172029	up	1.1384091	Aebp2
0.001870177	down	-1.2099588	Ci-Tyrp1/2a
0.040286023	down	-1.318311	RLBP1
0.036309548	down	-1.3236918	Blk
8.62E-04	down	-1.3940418	Ci-Tyrp1/2b
4.23E-05	down	-1.477952	ATPAF1
3.64E-05	down	-1.445624	OdsH
0.021763358	down	-1.462586	RAB3GAP1

0.001664294	down	-1.4881306	Ci-Tyrosinase
0.025531527	down	-1.520093	Ci-Noto2
4.23E-05	down	-1.556896	DEAF1
0.003077091	down	-1.5805709	CENTD1,RhoGAP15B
0.03313512	down	-1.5821983	ARRDC2
0.005342124	down	-1.6708257	CiGl
0.034525342	down	-1.6822704	FoxD3
0.009886694	down	-1.7366267	RHBDL1
5.44E-05	down	-1.8232851	Ci-Hunchbacklike,CASZ1
7.60E-04	down	-1.874163	TRPA1
2.86E-04	down	-1.9725444	Rfxdc1
0.017179336	down	-2.2694929	GJA10
0.0093702	down	-2.286389	Slc43a3
1.90E-04	down	-2.3616586	Pip5k3
2.50E-04	down	-2.5493562	Tcf-Lef
1.03E-05	down	-2.5672681	BMP5
4.64E-05	down	-2.6756895	GSC
4.47E-04	down	-2.6910844	Bzrap1
2.82E-04	down	-2.807287	aub
6.56E-04	down	-2.9526644	Mitf
3.63E-05	down	-3.6317787	SLC45A2

Table 2: Genes selected from the comparison dnFGFR and Control samples at 8 hpf			
p-value	regulation	fold change ratio	*Ciona* name
0.001672	up	2.833964	MLPH
0.027184	up	2.583631	BAPX1
3.36E-04	up	2.352395	ARL6
0.002847	up	1.946041	Slc7a9
0.049631	up	1.669716	GNRHR
0.038325	up	1.658639	Pax3/7
6.67E-04	up	1.603543	frizzled4
0.002078	up	1.597305	Rab38,RAB32,ltd
0.043638	up	1.585303	RGR
0.008557	up	1.545563	Macho1
0.001901	up	1.276154	Bcmo1
0.034862	up	1.243288	RLBP1
0.003237	up	1.217055	DOC2A
3.53E-04	up	1.225689	Myc
0.020556	up	1.174982	DHH,hedgehog homolog 1
4.17E-04	down	-1.228546	Mnt/Max
4.53E-04	down	-1.238756	Groucho

Table 3: Genes selected from the comparison EtsVp16 and Control samples at 12 hpf			
p-value	regulation	Fold change ratio	*Ciona* name
2.04E-05	up	4.334352	Blk

0.008398517	up	4.2540126	GJA10
3.66E-06	up	4.1178727	Mannose receptor-like 3
2.34E-04	up	4.066119	fringe
7.07E-04	up	3.7307255	RGA4
9.54E-05	up	3.411229	ELK3,Elk1
3.72E-04	up	3.2929611	Rfxdc1
0.001709635	up	2.684028	TRPA1
3.86E-05	up	2.4147341	ARRDC2
7.90E-04	up	2.375358	pcs,Sh3bp5
1.64E-04	up	2.2311928	Oxysterol binding protein-related protein 1
0.003217775	up	2.1357198	RHBDL1
3.33E-05	up	1.9871463	BMP5
0.011606308	up	1.8814628	Pip5k3
0.00428448	up	1.751554	CENTD1,RhoGAP15B
0.032720465	up	1.7431141	Mitf
0.004239461	up	1.7322165	Noto2
0.005612725	up	1.6993238	Dusp6.9
0.028443169	up	1.6009997	Frizzled receptor,fz
0.002715731	up	1.5422908	CBL
0.027991014	up	1.4788043	LRP4,Ci-Low-density lipoprotein receptor-related gene 1
0.010276981	up	1.4254256	NetA
0.006080867	up	1.4162877	SLC45A2
0.03816941	up	1.3764987	Piwi
0.040541194	up	1.3039634	Prickle
0.005235261	up	1.3022721	ARNT
0.005906015	up	1.2089932	Ci-Tyrp1/2a
0.015731048	up	1.1327978	Ci-Tyrp1/2b
0.01922549	up	1.092348	MXD1,Mnt
0.012398488	down	-1.3106596	Ci-macho1
0.024768367	down	-1.5276897	Ci-FoxD3
0.039910406	down	-1.594705	FoxB
0.020772489	down	-1.7842665	Pax6
0.001368281	down	-1.7849549	EVI1

REFERENCES

Aligianis, I. A., Johnson, C. A., Gissen, P., Chen, D., Hampshire, D., Hoffmann, K., Maina, E. N., Morgan, N. V., Tee, L., Morton, J. et al. (2005). Mutations of the catalytic subunit of RAB3GAP cause Warburg Micro syndrome. *Nat Genet* **37**, 221-3.

Alto, N. M. (2002). Rab32 is an A-kinase anchoring protein and participates in mitochondrial dynamics. *The Journal of Cell Biology* **158**, 659-668.

Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I. and Land, H. (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* **359**, 423-6.

Bansal, M., Belcastro, V., Ambesi-Impiombato, A., di Bernardo, D. (2007)**.** How to infer gene networks from expression profiles. *Molecular Systems Biology* 3:78

Beaster-Jones, L., Kaltenbach, S. L., Koop, D., Yuan, S., Chastain, R. and Holland, L. Z. (2008). Expression of somite segmentation genes in amphioxus: a clock without a wavefront? *Dev Genes Evol* **218**, 599-611.

Beh, J., Shi, W., Levine, M., Davidson, B. and Christiaen, L. (2007). FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian Ciona intestinalis. *Development* **134**, 3297-305.

Bennett, D. C. and Lamoreux, M. L. (2003). The color loci of mice--a genetic century. *Pigment Cell Res* **16**, 333-44.

Berson, J. F., Harper, D. C., Tenza, D., Raposo, G. and Marks, M. S. (2001). Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. *Mol Biol Cell* **12**, 3451-64.

Bertrand, V., Hudson, C., Caillol, D., Popovici, C. and Lemaire, P. (2003). Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell* **115**, 615-27.

Bolstad, B. M., Irizarry, R. A., Astrand, M. and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-93.

Bonifacino, J. S. (2004). Insights into the biogenesis of lysosome-related organelles from the study of the Hermansky-Pudlak syndrome. *Ann N Y Acad Sci* **1038**, 103-14.

Bouchard, B., Fuller, B. B., Vijayasaradhi, S. and Houghton, A. N. (1989). Induction of pigmentation in mouse fibroblasts by expression of human tyrosinase cDNA. *J Exp Med* **169**, 2029-42.

Busch, W. and Lohmann, J. U. (2007). Profiling a plant: expression analysis in Arabidopsis. *Curr Opin Plant Biol* **10**, 136-41.

Cajal, M., Lawson, K. A., Hill, B., Moreau, A., Rao, J., Ross, A., Collignon, J. and Camus, A. (2012). Clonal and molecular analysis of the prospective anterior neural boundary in the mouse embryo. *Development* **139**, 423-36.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

Caracciolo, A., Gesualdo, I., Branno, M., Aniello, F., Di Lauro, R. and Palumbo, A. (1997). Specific cellular localization of tyrosinase mRNA during Ciona intestinalis larval development. *Dev Growth Differ* **39**, 437-44.

Choi, H. Y., Dieckmann, M., Herz, J. and Niemeier, A. (2009). Lrp4, a novel receptor for Dickkopf 1 and sclerostin, is expressed by osteoblasts and regulates bone growth and turnover in vivo. *PLoS ONE* **4**, e7930.

Christiaen, L., Davidson, B., Kawashima, T., Powell, W., Nolla, H., Vranizan, K. and Levine, M. (2008). The Transcription/Migration Interface in Heart Precursors of Ciona intestinalis. *Science* **320**, 1349-1352.

Claus, H. and Decker, H. (2006). Bacterial tyrosinases. *Syst Appl Microbiol* **29**, 3-14.

Corbo, J. C., Erives, A., Di Gregorio, A., Chang, A. and Levine, M. (1997). Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. *Development* **124**, 2335-44.

Costin, G. E., Valencia, J. C., Vieira, W. D., Lamoreux, M. L. and Hearing, V. J. (2003). Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4. *J Cell Sci* **116**, 3203-12.

Cress, W. D. and Triezenberg, S. J. (1991). Critical structural elements of the VP16 transcriptional activation domain. *Science* **251**, 87-90.

Cui, Y. and Freedman, J. H. (2009). Cadmium induces retinoic acid signaling by regulating retinoic acid metabolic gene expression. *J Biol Chem* **284**, 24925-32.

D'Adamo, P., Menegon, A., Lo Nigro, C., Grasso, M., Gulisano, M., Tamanini, F., Bienvenu, T., Gedeon, A. K., Oostra, B., Wu, S. K. et al. (1998). Mutations in GDI1 are responsible for X-linked non-specific mental retardation. *Nat Genet* **19**, 134-9.

Dahlquist, K. D., Salomonis, N., Vranizan, K., Lawlor, S. C. and Conklin, B. R. (2002). GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* **31**, 19-20.

Davidson, B., Shi, W., Beh, J., Christiaen, L. and Levine, M. (2006). FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. *Genes Dev* **20**, 2728-38.

De Cegli R, Iacobacci S, Flore G, Gambardella G, Mao L, Cutillo L, Lauria M, Klose J, Illingworth E, Banfi S, di Bernardo D. (2013). Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. *Nucleic Acids Res. 2013 Jan;41(2):711-26*

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M. et al. (2002). The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science* **298**, 2157-67.

del Marmol, V. and Beermann, F. (1996). Tyrosinase and related proteins in mammalian pigmentation. *FEBS Lett* **381**, 165-8.

Delsuc, F., Brinkmann, H., Chourrout, D. and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965-8.

Di Gregorio, A. and Levine, M. (2002). Analyzing gene regulation in ascidian embryos: new tools for new perspectives. *Differentiation* **70**, 132-9.

Di Pietro, S. M. and Dell'Angelica, E. C. (2005). The cell biology of Hermansky-Pudlak syndrome: recent advances. *Traffic* **6**, 525-33.

Dilly, P. N. (1969). Studies on the receptors in Ciona intestinalis. 3. A second type of photoreceptor in the tadpole larva of Ciona intestinalis. *Z Zellforsch Mikrosk Anat* **96**, 63-5.

Dono, R. (2003). Fibroblast growth factors as regulators of central nervous system development and function. *Am J Physiol Regul Integr Comp Physiol* **284**, R867-81.

Donoghue, P. C. and Purnell, M. A. (2005). Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol* **20**, 312-9.

Eakin, R. M. (1973). The third eye. University of California Press.

Eakin, R. M. and Kuda, A. (1971). Ultrastructure of sensory receptors in Ascidian tadpoles. *Z Zellforsch Mikrosk Anat* **112**, 287-312.

Esposito, R., D'Aniello, S., Squarzoni, P., Pezzotti, M. R., Ristoratore, F. and Spagnuolo, A. (2012). New insights into the evolution of metazoan tyrosinase gene family. *PLoS ONE* **7**, e35731.

Faith J, Gardner T (2005) Reverse-engineering transcription control networks. *Phys Life Rev* 2: 65–88

Farnsworth, C. L. and Feig, L. A. (1991). Dominant inhibitory mutations in the Mg(2+)-binding site of RasH prevent its activation by GTP. *Mol Cell Biol* **11**, 4822-9.

Fisher, A. L., Ohsako, S. and Caudy, M. (1996). The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. *Mol Cell Biol* **16**, 2670-7.

Fujikawa, K., Satoh, A. K., Kawamura, S. and Ozaki, K. (2002a). Molecular and Functional Characterization of a Unique Rab Protein, RABRP1, Containing the WDIAGQE Sequence in a GTPase Motif. *Zoological Science* **19**, 981-993.

Fujikawa, K., Satoh, A. K., Kawamura, S. and Ozaki, K. (2002b). Molecular and functional characterization of a unique Rab protein, RABRP1, containing the WDIAGQE sequence in a GTPase motif. *Zoolog Sci* **19**, 981-93.

Fukuda, M. (2005). Versatile Role of Rab27 in Membrane Trafficking: Focus on the Rab27 Effector Families. *Journal of Biochemistry* **137**, 9-16.

Giorgi, F. M., Bolger, A. M., Lohse, M. and Usadel, B. (2010). Algorithm-driven artifacts in median Polish summarization of microarray data. *BMC Bioinformatics* **11**, 553.

Goding, C. (2007). Melanoma, making a difference. *Pigment Cell Res* **20**, 457.

Grandori, C., Cowley, S. M., James, L. P. and Eisenman, R. N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol* **16**, 653-99.

Grosshans, B. L., Ortiz, D. and Novick, P. (2006). Rabs and their effectors: achieving specificity in membrane traffic. *Proc Natl Acad Sci U S A* **103**, 11821-7.

Haeussler, M., Jaszczyszyn, Y., Christiaen, L. and Joly, J. S. (2010). A cis-regulatory signature for chordate anterior neuroectodermal genes. *PLoS Genet* **6**, e1000912.

Hall, D. B. and Struhl, K. (2002). The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *J Biol Chem* **277**, 46043-50.

Ho Sui, S. J., Fulton, D. L., Arenillas, D. J., Kwon, A. T. and Wasserman, W. W. (2007). oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res* **35**, W245-52.

Hollenhorst, P. C., Ferris, M. W., Hull, M. A., Chae, H., Kim, S. and Graves, B. J. (2011). Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-57.

Horie, T., Orii, H. and Nakagawa, M. (2005). Structure of ocellus photoreceptors in the ascidian Ciona intestinalis larva as revealed by an anti-arrestin antibody. *J Neurobiol* **65**, 241-50.

Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T. and Ikeo, K. (2007). A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev Dyn* **236**, 1790-805.

Hudson, C., Lotito, S. and Yasuo, H. (2007a). Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. *Development* **134**, 3527-3537.

Hudson, C., Lotito, S. and Yasuo, H. (2007b). Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. *Development* **134**, 3527-37.

Hume, A. N., Tarafder, A. K., Ramalho, J. S., Sviderskaya, E. V. and Seabra, M. C. (2006). A coiled-coil domain of melanophilin is essential for Myosin Va recruitment and melanosome transport in melanocytes. *Mol Biol Cell* **17**, 4720-35.

Imai, K. S. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* **131**, 4047-4058.

Imai, K. S., Hino, K., Yagi, K., Satoh, N. and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* **131**, 4047-58.

Imai, K. S., Levine, M., Satoh, N. and Satou, Y. (2006). Regulatory blueprint for a chordate embryo. *Science* **312**, 1183-7.

Imai, K. S., Stolfi, A., Levine, M. and Satou, Y. (2009). Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. *Development* **136**, 285-93.

Irizarry, R. A., Cope, L. M. and Wu, Z. (2006). Feature-level exploration of a published Affymetrix GeneChip control dataset. *Genome Biol* **7**, 404.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64.

Jeffery, G. (2001). Architecture of the optic chiasm and the mechanisms that sculpt its development. *Physiol Rev* **81**, 1393-414.

Jeffery, W. R. and Meier, S. (1983). A yellow crescent cytoskeletal domain in ascidian eggs and its role in early development. *Dev Biol* **96**, 125-43.

Jiao, R., Daube, M., Duan, H., Zou, Y., Frei, E. and Noll, M. (2001). Headless flies generated by developmental pathway interference. *Development* **128**, 3307-19.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**, D277-80.

Kang, S. A., Seol, J. H. and Kim, J. (2005). The conserved WRPW motif of Hes6 mediates proteasomal degradation. *Biochem Biophys Res Commun* **332**, 33-6.

Kasprzyk, A. (2011). BioMart: driving a paradigm change in biological data management. *Database (Oxford)* **2011**, bar049.

Kendall, M. G., Stuart, A., Ord, K., Arnold, S. and O'Hagan, A. (1994). Kendall's Advanced Theory of Statistics, Classical Inference and the Linear Model: John Wiley & Sons.

King-Smith, C. and Cronin, T. W. (1996). Pigment granule migration in crustacean photoreceptors requires calcium. *Vis Neurosci* **13**, 43-9.

Kukimoto-Niino, M., Sakamoto, A., Kanno, E., Hanawa-Suetsugu, K., Terada, T., Shirouzu, M., Fukuda, M. and Yokoyama, S. (2008). Structural basis for the exclusive specificity of Slac2-a/melanophilin for the Rab27 GTPases. *Structure* **16**, 1478-90.

Kusakabe, T., Kusakabe, R., Kawakami, I., Satou, Y., Satoh, N. and Tsuda, M. (2001). Ci-opsin1, a vertebrate-type opsin gene, expressed in the larval ocellus of the ascidian Ciona intestinalis. *FEBS Lett* **506**, 69-72.

Kwon, A. T., Arenillas, D. J., Worsley Hunt, R. and Wasserman, W. W. (2012). oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3 (Bethesda)* **2**, 987-1002.

Lagutin, O. V. (2003). Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes & Development* **17**, 368-379.

Lamb, T. D., Collin, S. P. and Pugh, E. N., Jr. (2007). Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nat Rev Neurosci* **8**, 960-76.

LeDouarin. (1982). The neural crest. Cambridge: Cambridge University Press.

Lee, L. P., Lai, M. H., Chiu, W. K., Leung, M. W., Liu, K. K. and Chan, H. B. (2010). Management of primary spontaneous pneumothorax in Chinese children. *Hong Kong Med J* **16**, 94-100.

Lee, M. T., Mishra, A. and Lambright, D. G. (2009). Structural mechanisms for regulation of membrane traffic by rab GTPases. *Traffic* **10**, 1377-89.

Loftus, S. K. (2002). Mutation of melanosome protein RAB38 in chocolate mice. *Proceedings of the National Academy of Sciences* **99**, 4471-4476.

Loftus, S. K., Larson, D. M., Baxter, L. L., Antonellis, A., Chen, Y., Wu, X., Jiang, Y., Bittner, M., Hammer, J. A., 3rd and Pavan, W. J. (2002). Mutation of melanosome protein RAB38 in chocolate mice. *Proc Natl Acad Sci U S A* **99**, 4471-6.

Loo, L. W., Secombe, J., Little, J. T., Carlos, L. S., Yost, C., Cheng, P. F., Flynn, E. M., Edgar, B. A. and Eisenman, R. N. (2005). The transcriptional repressor dMnt is a regulator of growth in Drosophila melanogaster. *Mol Cell Biol* **25**, 7078-91.

Lopes, V. S., Wasmeier, C., Seabra, M. C. and Futter, C. E. (2007). Melanosome Maturation Defect in Rab38-deficient Retinal Pigment Epithelium Results in Instability of Immature Melanosomes during Transient Melanogenesis. *Molecular Biology of the Cell* **18**, 3914-3927.

Ma, J. (2004). Lightoid and Claret: A rab GTPase and its putative guanine nucleotide exchange factor in biogenesis of Drosophila eye pigment granules. *Proceedings of the National Academy of Sciences* **101**, 11652-11657.

Marks, M. S. and Seabra, M. C. (2001). The melanosome: membrane dynamics in black and white. *Nat Rev Mol Cell Biol* **2**, 738-48.

Matesic, L. E., Yip, R., Reuss, A. E., Swing, D. A., O'Sullivan, T. N., Fletcher, C. F., Copeland, N. G. and Jenkins, N. A. (2001). Mutations in Mlph, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice. *Proc Natl Acad Sci U S A* **98**, 10238-43.

Matsui, Y., Kikuchi, A., Araki, S., Hata, Y., Kondo, J., Teranishi, Y. and Takai, Y. (1990). Molecular cloning and characterization of a novel type of regulatory protein (GDI) for smg p25A, a ras p21-like GTP-binding protein. *Mol Cell Biol* **10**, 4116-22.

Maxfield, F. R. and Mondal, M. (2006). Sterol and lipid trafficking in mammalian cells. *Biochem Soc Trans* **34**, 335-9.

Mazet, F., Hutt, J. A., Millard, J. and Shimeld, S. M. (2003). Pax gene expression in the developing central nervous system of Ciona intestinalis. *Gene Expr Patterns* **3**, 743-5.

Mazet, F., Hutt, J. A., Milloz, J., Millard, J., Graham, A. and Shimeld, S. M. (2005). Molecular evidence from Ciona intestinalis for the evolutionary origin of vertebrate sensory placodes. *Dev Biol* **282**, 494-508.

Meinertzhagen, I. A. and Okamura, Y. (2001). The larval ascidian nervous system: the chordate brain from its small beginnings. *Trends Neurosci* **24**, 401-10.

Meredith, P. and Sarna, T. (2006). The physical and chemical properties of eumelanin. *Pigment Cell Res* **19**, 572-94.

Millenaar, F. F., Okyere, J., May, S. T., van Zanten, M., Voesenek, L. A. and Peeters, A. J. (2006). How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics* **7**, 137.

Mittelstaedt, T. and Schoch, S. (2007). Structure and evolution of RIM-BP genes: identification of a novel family member. *Gene* **403**, 70-9.

Moret, F., Christiaen, L., Deyts, C., Blin, M., Vernier, P. and Joly, J. S. (2005). Regulatory gene expressions in the ascidian ventral sensory vesicle: evolutionary relationships with the vertebrate hypothalamus. *Dev Biol* **277**, 567-79.

Nakagawa, M., Orii, H., Yoshida, N., Jojima, E., Horie, T., Yoshida, R., Haga, T. and Tsuda, M. (2002). Ascidian arrestin (Ci-arr), the origin of the visual and nonvisual arrestins of vertebrate. *Eur J Biochem* **269**, 5112-8.

Nakashima, Y., Kusakabe, T., Kusakabe, R., Terakita, A., Shichida, Y. and Tsuda, M. (2003). Origin of the vertebrate visual cycle: genes encoding retinal photoisomerase and two putative visual cycle proteins are expressed in whole brain of a primitive chordate. *J Comp Neurol* **460**, 180-90.

Newton, J. M., Cohen-Barak, O., Hagiwara, N., Gardner, J. M., Davisson, M. T., King, R. A. and Brilliant, M. H. (2001). Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. *Am J Hum Genet* **69**, 981-8.

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev Biol* **121**, 526-41.

Nishida, H. and Satoh, N. (1989). Determination and regulation in the pigment cell lineage of the ascidian embryo. *Dev Biol* **132**, 355-67.

Nuoffer, C., Davidson, H. W., Matteson, J., Meinkoth, J. and Balch, W. E. (1994). A GDP-bound of rab1 inhibits protein export from the endoplasmic reticulum and transport between Golgi compartments. *J Cell Biol* **125**, 225-37.

Ogasawara, M., Di Lauro, R. and Satoh, N. (1999). Ascidian homologs of mammalian thyroid peroxidase genes are expressed in the thyroid-equivalent region of the endostyle. *J Exp Zool* **285**, 158-69.

Ohazama, A., Johnson, E. B., Ota, M. S., Choi, H. Y., Porntaveetus, T., Oommen, S., Itoh, N., Eto, K., Gritli-Linde, A., Herz, J. et al. (2008). Lrp4 modulates extracellular integration of cell signaling pathways in development. *PLoS ONE* **3**, e4092.

Ohbayashi, N. and Fukuda, M. (2012). Role of Rab family GTPases and their effectors in melanosomal logistics. *J Biochem* **151**, 343-51.

Ohtsuki, H. (1991). Sensory organs in the cerebral vesicle of the ascidian larva, Aplidium sp.: an SEM study. *Zoolog Sci* **8**, 235-242.

Oikawa, T. and Yamada, T. (2003). Molecular biology of the Ets family of transcription factors. *Gene* **303**, 11-34.

Oiso, N., Riddle, S. R., Serikawa, T., Kuramoto, T. and Spritz, R. A. (2004). The rat Ruby (R) locus is Rab38: identical mutations in Fawn-hooded and Tester-Moriyama rats derived from an ancestral Long Evans rat sub-strain. *Mamm Genome* **15**, 307-14.

Olivares, C. and Solano, F. (2009). New insights into the active site structure and catalytic mechanism of tyrosinase and its related proteins. *Pigment Cell Melanoma Res* **22**, 750-60.

Orlow, S. J. (1995). Congenital disorders of hypopigmentation. *Semin Dermatol* **14**, 27-32.

Palumbo, A., d'Ischia, M., Misuraca, G. and Prota, G. (1991). Mechanism of inhibition of melanogenesis by hydroquinone. *Biochim Biophys Acta* **1073**, 85-90.

Park, C. S., Lee, S. W., Kim, Y. S., Kim, E. J., Sin, H. S., Oh, D. K., Kim, S. W. and Um, S. J. (2008). Utilization of the recombinant human beta-carotene-15,15'-monooxygenase gene in Escherichia coli and mammalian cells. *Biotechnol Lett* **30**, 735-41.

Pereira-Leal, J. B. and Seabra, M. C. (2001). Evolution of the Rab family of small GTP-binding proteins. *J Mol Biol* **313**, 889-901.

Pfeffer, S. R. (2001). Rab GTPases: specifying and deciphering organelle identity and function. *Trends Cell Biol* **11**, 487-91.

Pfeffer, S. R. (2005). Structural clues to Rab GTPase functional diversity. *J Biol Chem* **280**, 15485-8.

Plonka, P. M. and Grabacka, M. (2006). Melanin synthesis in microorganisms--biotechnological and medical aspects. *Acta Biochim Pol* **53**, 429-43.

Raposo, G. and Marks, M. S. (2007). Melanosomes — dark organelles enlighten endosomal membrane transport. *Nature Reviews Molecular Cell Biology* **8**, 786-797.

Raposo, G., Tenza, D., Murphy, D. M., Berson, J. F. and Marks, M. S. (2001). Distinct protein sorting and localization to premelanosomes, melanosomes, and lysosomes in pigmented melanocytic cells. *J Cell Biol* **152**, 809-24.

Ribisi, S., Jr., Mariani, F. V., Aamar, E., Lamb, T. M., Frank, D. and Harland, R. M. (2000). Ras-mediated FGF signaling is required for the formation of posterior but not anterior neural tissue in Xenopus laevis. *Dev Biol* **227**, 183-96.

Riley, P. A. (1992). Materia melanica: further dark thoughts. *Pigment Cell Res* **5**, 101-6.

Sanchez-Arrones, L., Stern, C. D., Bovolenta, P. and Puelles, L. (2012). Sharpening of the anterior neural border in the chick by rostral endoderm signalling. *Development* **139**, 1034-44.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4.

Sarna. (1996). Properties and function of the ocular melanin--a photobiophysical view.

Sato, M., Mori, Y., Matsui, T., Aoki, R., Oya, M., Yanagihara, Y., Fukuda, M. and Tsuboi, T. (2010). Role of the polybasic sequence in the Doc2alpha C2B domain in dense-core vesicle exocytosis in PC12 cells. *J Neurochem* **114**, 171-81.

Sato, S. and Yamamoto, H. (2001). Development of pigment cells in the brain of ascidian tadpole larvae: insights into the origins of vertebrate pigment cells. *Pigment Cell Res* **14**, 428-36.

Satoh, N. (2001). Ascidian embryos as a model system to analyze expression and function of developmental genes. *Differentiation* **68**, 1-12.

Satou, Y., Hamaguchi, M., Takeuchi, K., Hastings, K. E. and Satoh, N. (2006). Genomic overview of mRNA 5'-leader trans-splicing in the ascidian Ciona intestinalis. *Nucleic Acids Res* **34**, 3378-88.

Satou, Y., Imai, K. S., Levine, M., Kohara, Y., Rokhsar, D. and Satoh, N. (2003a). A genomewide survey of developmentally relevant genes in Ciona intestinalis. I. Genes for bHLH transcription factors. *Dev Genes Evol* **213**, 213-21.

Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. and Satoh, N. (2005). An integrated database of the ascidian, Ciona intestinalis: towards functional genomics. *Zoolog Sci* **22**, 837-43.

Satou, Y., Sasakura, Y., Yamada, L., Imai, K. S., Satoh, N. and Degnan, B. (2003b). A genomewide survey of developmentally relevant genes in Ciona intestinalis. V. Genes for receptor tyrosine kinase pathway and Notch signaling pathway. *Dev Genes Evol* **213**, 254-63.

Satou, Y., Yagi, K., Imai, K. S., Yamada, L., Nishida, H. and Satoh, N. (2002). macho-1-Related genes in Ciona embryos. *Dev Genes Evol* **212**, 87-92.

Schubert, M., Escriva, H., Xavier-Neto, J. and Laudet, V. (2006). Amphioxus and tunicates as evolutionary model systems. *Trends Ecol Evol* **21**, 269-77.

Shirane, M. and Nakayama, K. I. (2006). Protrudin induces neurite formation by directional membrane trafficking. *Science* **314**, 818-21.

Sichel, G., Corsaro, C., Scalia, M., Sciuto, S. and Geremia, E. (1987). Relationship between melanin content and superoxide dismutase (SOD) activity in the liver of various species of animals. *Cell Biochem Funct* **5**, 123-8.

Silvers. (1979). The coat colors of mice - A model for mammalian gene action and interaction. New York: Springer-Verlag.

Spritz, R. A., Chiang, P. W., Oiso, N. and Alkhateeb, A. (2003). Human and mouse disorders of pigmentation. *Curr Opin Genet Dev* **13**, 284-9.

Squarzoni, P., Parveen, F., Zanetti, L., Ristoratore, F. and Spagnuolo, A. (2011). FGF/MAPK/Ets signaling renders pigment cell precursors competent to respond to Wnt signal by directly controlling Ci-Tcf transcription. *Development* **138**, 1421-1432.

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H. et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12**, 1611-8.

Steel, K. P. and Barkway, C. (1989). Another role for melanocytes: their importance for normal stria vascularis development in the mammalian inner ear. *Development* **107**, 453-63.

Steel, K. P., Davidson, D. R. and Jackson, I. J. (1992). TRP-2/DT, a new early melanoblast marker, shows that steel growth factor (c-kit ligand) is a survival factor. *Development* **115**, 1111-9.

Stolfi, A., Wagner, E., Taliaferro, J. M., Chou, S. and Levine, M. (2011). Neural tube patterning by Ephrin, FGF and Notch signaling relays. *Development* **138**, 5429-39.

Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. and Stern, C. D. (2000). Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406**, 74-8.

Sulaimon, S. S. and Kitchell, B. E. (2003). The biology of melanocytes. *Vet Dermatol* **14**, 57-65.

Szebenyi, G. and Fallon, J. F. (1999). Fibroblast growth factors as multifunctional signaling factors. *Int Rev Cytol* **185**, 45-106.

Takano, Y., Kubo, Y., Kawamura, C., Tsuge, T. and Furusawa, I. I. (1997). The Alternaria alternata Melanin Biosynthesis Gene Restores Appressorial Melanization and Penetration of Cellulose Membranes in the Melanin-Deficient Albino Mutant of Colletotrichum lagenarium. *Fungal Genet Biol* **21**, 131-40.

Tamura, K., Ohbayashi, N., Ishibashi, K. and Fukuda, M. (2011). Structure-function analysis of VPS9-ankyrin-repeat protein (Varp) in the trafficking of tyrosinase-related protein 1 in melanocytes. *J Biol Chem* **286**, 7507-21.

Tamura, K., Ohbayashi, N., Maruta, Y., Kanno, E., Itoh, T. and Fukuda, M. (2009). Varp is a novel Rab32/38-binding protein that regulates Tyrp1 trafficking in melanocytes. *Mol Biol Cell* **20**, 2900-8.

Tassy, O., Daian, F., Hudson, C., Bertrand, V. and Lemaire, P. (2006). A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. *Curr Biol* **16**, 345-58.

Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R. et al. (2010). The ANISEED database: Digital representation, formalization, and elucidation of a chordate developmental program. *Genome Research* **20**, 1459-1468.

Tief, K., Hahne, M., Schmidt, A. and Beermann, F. (1996). Tyrosinase, the key enzyme in melanin synthesis, is expressed in murine brain. *Eur J Biochem* **241**, 12-6.

Toro, S. and Varga, Z. M. (2007). Equivalent progenitor cells in the zebrafish anterior preplacodal field give rise to adenohypophysis, lens, and olfactory placodes. *Semin Cell Dev Biol* **18**, 534-42.

Torrence. (1986). Sensory endings of the ascidian static organ (Chordata, Ascidiacea). *Zoomorphology* **106**, 61-66.

Toyoda, R., Kasai, A., Sato, S., Wada, S., Saiga, H., Ikeo, K., Gojobori, T., Numakunai, T. and Yamamoto, H. (2004). Pigment cell lineage-specific expression activity of the ascidian tyrosinase-related gene. *Gene* **332**, 61-9.

Tsuda, M., Kawakami, I. and Shiraishi, S. (2003a). Sensitization and habituation of the swimming behavior in ascidian larvae to light. *Zoolog Sci* **20**, 13-22.

Tsuda, M., Sakurai, D. and Goda, M. (2003b). Direct evidence for the role of pigment cells in the brain of ascidian larvae by laser ablation. *J Exp Biol* **206**, 1409-17.

Tukey, J. W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science* **198**, 679-84.

Van Gele, M., Dynoodt, P. and Lambert, J. (2009). Griscelli syndrome: a model system to study vesicular trafficking. *Pigment Cell Melanoma Res* **22**, 268-82.

Vasiliauskas, D. and Stern, C. D. (2001). Patterning the embryonic axis: FGF signaling and how vertebrate embryos measure time. *Cell* **106**, 133-6.

Vetter, I. R. and Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299-304.

Vienne, A. and Pontarotti, P. (2006). Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int J Biol Sci* **2**, 32-7.

Wada, H., Saiga, H., Satoh, N. and Holland, P. W. (1998). Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian Pax-2/5/8, Hox and Otx genes. *Development* **125**, 1113-22.

Wada, S., Tokuoka, M., Shoguchi, E., Kobayashi, K., Di Gregorio, A., Spagnuolo, A., Branno, M., Kohara, Y., Rokhsar, D., Levine, M. et al. (2003). A genomewide survey

of developmentally relevant genes in Ciona intestinalis. II. Genes for homeobox transcription factors. *Dev Genes Evol* **213**, 222-34.

Wagner, E. and Levine, M. (2012). FGF signaling establishes the anterior border of the Ciona neural tube. *Development* **139**, 2351-9.

Wakamatsu, K., Ito, S. and Rees, J. L. (2002). The usefulness of 4-amino-3-hydroxyphenylalanine as a specific marker of pheomelanin. *Pigment Cell Res* **15**, 225-32.

Wang, Y., Okamoto, M., Schmitz, F., Hofmann, K. and Sudhof, T. C. (1997). Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion. *Nature* **388**, 593-8.

Wasmeier, C., Romao, M., Plowright, L., Bennett, D. C., Raposo, G. and Seabra, M. C. (2006). Rab38 and Rab32 control post-Golgi trafficking of melanogenic enzymes. *J Cell Biol* **175**, 271-81.

Wei, M. L. (2006). Hermansky-Pudlak syndrome: a disease of protein trafficking and organelle function. *Pigment Cell Res* **19**, 19-42.

Whittaker, J. R. (1973). Tyrosinase in the presumptive pigment cells of ascidian embryos: tyrosine accessibility may initiate melanin synthesis. *Dev Biol* **30**, 441-54.

Wilson, S. I., Graziano, E., Harland, R., Jessell, T. M. and Edlund, T. (2000). An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo. *Curr Biol* **10**, 421-9.

Wittkopp, P. J., Vaccaro, K. and Carroll, S. B. (2002). Evolution of yellow gene regulation and pigmentation in Drosophila. *Curr Biol* **12**, 1547-56.

Wu, X., Vasisht, V., Kosman, D., Reinitz, J. and Small, S. (2001). Thoracic patterning by the Drosophila gap gene hunchback. *Dev Biol* **237**, 79-92.

Wu, X. S., Rao, K., Zhang, H., Wang, F., Sellers, J. R., Matesic, L. E., Copeland, N. G., Jenkins, N. A. and Hammer, J. A., 3rd. (2002). Identification of an organelle receptor for myosin-Va. *Nat Cell Biol* **4**, 271-8.

Wu, Z. and Irizarry, R. A. (2004). Preprocessing of oligonucleotide array data. *Nat Biotechnol* **22**, 656-8; author reply 658.

Yang, S. H., Bumpass, D. C., Perkins, N. D. and Sharrocks, A. D. (2002). The ETS domain transcription factor Elk-1 contains a novel class of repression domain. *Mol Cell Biol* **22**, 5036-46.

Zalokar, M. and Sardet, C. (1984). Tracing of cell lineage in embryonic development of Phallusia mammillata (Ascidia) by vital staining of mitochondria. *Dev Biol* **102**, 195-205.

Zang, J., Sun, Y., Wang, Y., Yang, J., Li, F., Zhou, Y., Zhu, L., Jessica, R., Mohammadhosein, F., Xu, J. et al. (2008). Dissection of genetic overlap of salt tolerance QTLs at the seedling and tillering stages using backcross introgression lines in rice. *Sci China C Life Sci* **51**, 583-91.

Zhou, Z. Q. and Hurlin, P. J. (2001). The interplay between Mad and Myc in proliferation and differentiation. *Trends Cell Biol* **11**, S10-4.

Zhu, C. C., Dyer, M. A., Uchikawa, M., Kondoh, H., Lagutin, O. V. and Oliver, G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. *Development* **129**, 2835-49.